QUARTERLY OF APPLIED MATHEMATICS

Vol. V JANUARY, 1948 No. 4

RELATIONS BETWEEN THE TRANSMITTING AND RECEIVING PROPERTIES OF ANTENNAS*

BY

A. F. STEVENSON

University of Toronto

1. Introduction. It has been known for some time that there exist relations between the transmitting and receiving properties of antennas. Such a relation was first established by Carson,[1] and results of like character have been given by Ballantine[2] and Sommerfeld.[3] Mention may also be made of discussions by Slater[4] and Schelkunoff.[5]

When critically examined, however, both the statement and the proof of the results claimed leave much to be desired from the point of view of preciseness and rigour. Without attempting a detailed criticism of previous discussions of the subject, it appears to the writer that such discussions are all open to at least some of the following objections.

(1) Results proved for *circuits*—e.g. Thévenin's theorem—are assumed without proof to apply to *antennas*.

(2) The boundary conditions to be satisfied at the antenna by the field quantities are not precisely stated.

(3) No distinction is made between an antenna acting as *receiver*, with connected load impedance, and one acting as a *parasite*.

(4) The "impressed voltage" is introduced in an unsatisfactory manner which is not consistent with Maxwell's equations. In particular, no account is taken of the "gap."

(5) "Reciprocal theorems" proved for two antennas are assumed to establish results relating to the transmitting and receiving properties for a *single* antenna. Clearly, it should be possible to establish such results without referring to a second antenna at all.

(6) The antenna is treated as an infinitesimal Hertzian dipole, which is an unnecessary and unwarranted simplification (this applies to Sommerfeld's[3] work).

In this paper, a discussion of the subject is presented which, it is believed, is more thorough and rigorous than any hitherto given. In Sec. 2 a rigorous formulation, in

* Received Dec. 26, 1946.

[1] J. R. Carson, Bell System Tech. J., **3**, 393 (1924); **9**, 325 (1930); Proc. I.R.E., **17**, 952 (1929).

[2] S. Ballantine, Proc. I.R.E., **17**, 929 (1929).

[3] A. Sommerfeld, Z. für Hochfrequenztechnik, **26**, 93 (1925). See also the article by Sommerfeld in Frank-Mises, *Differentialgleichungen der Physik*, Rosenberg Publ. N.Y.C., 1943, vol. II, 7th ed., p. 576.

[4] J. C. Slater, *Microwave Transmission*, McGraw-Hill, 1942, chap. 6.

[5] S. A. Schelkunoff, *Electromagnetic Waves*, Van Nostrand 1943, p. 476.

terms of field theory, of three fundamental mathematical problems associated with an antenna, namely those of a transmitting, parasitic, and receiving antenna, is given, it being assumed that the antenna is perfectly conducting. The surrounding medium may be of a very general (linear) type. Attention is confined throughout to a steady state with a definite frequency. In Sec. 3 it is shown that of the three problems mentioned, one, namely that of reception, can be solved very simply in terms of the solutions of the other two. Incidentally, a rigorous proof of Thévenin's theorem for antennas, without any appeal to circuit theory, is given. In Sec. 4 it is shown further that there is a connection between the transmission and parasite problems which, effectively, enables us to solve the latter problem (and hence also the reception problem) once the former has been solved. An application is made to the case where the whole medium is homogeneous and isotropic, and in particular to a linear antenna in such a medium.

In Sec. 5 reciprocal relations between two antennas are discussed. These are of the usual type, though it is believed that they are somewhat more precisely stated than is usually the case. An apparent contradiction of the reciprocal relation which has been put forward by Ballantine[2] (originally in criticism of Sommerfeld's[3] work) is here briefly dealt with. In Sec. 6 the question is discussed as to whether the directional properties of an antenna, as exhibited by a polar diagram, are identical for transmission and reception. Adopting a reasonable definition of "polar diagram," it is concluded that the identity of the polar diagrams does not hold in general, although it does so (at least approximately) in cases which are likely to occur in practice. In Sec. 7 the results are extended to cover the case where the antenna is imperfectly conducting. In an appendix a theorem of Lorentz, which is used in the course of the work, is extended to the case where the medium is anisotropic.

The results of Secs. 4 and 5, and subsequent results based on these, involve a certain approximation; but the errors involved would be negligible in practical cases.

2. The three mathematical problems associated with an antenna. We shall consider the antenna to be perfectly conducting, though this restriction will be removed in Sec. 7. The surrounding medium will be supposed to be characterized by linear constitutive relations so that the field equations are all linear. These relations need not, however, unless the contrary is stated, be of the usual isotropic type. We may suppose that the medium is crystalline and is characterized by symmetric dielectric, permeability and conductivity tensors ϵ_{ij}, μ_{ij}, σ_{ij} whose components are arbitrary functions of position. Further, these tensors may have arbitrary surfaces of discontinuity.[6] We shall confine ourselves to a steady state solution in which all field quantities vary with the time according to the factor $e^{-i\omega t}$, this factor being omitted, and we shall employ Gaussian units throughout.

We shall, for the sake of generality, consider an "antenna" to be any closed perfectly conducting surface; usually this surface will, of course, take the form of a wire or system of wires. Since the antenna is perfectly conducting, the tangential components of **E** vanish everywhere on the antenna surface except where there is an impressed (or induced) E.M.F. or "voltage." Such voltages occur when the antenna is acting as a transmitter or receiver, but not, of course, when it is acting as a parasite. To treat the case where there is a voltage in a reasonably simple manner which en-

[6] There may also be perfect conductors in the surrounding medium ($\sigma = \infty$). This case can either be treated as a limiting case or considered *ab initio*.

ables us to dispense with the connecting circuit (as is always done in such cases), we shall ignore the connecting circuit, but suppose that there is a small "gap" on the antenna surface where the transmitting or receiving circuit is connected.[7] The component of **E** across this gap does *not* vanish, but rises to a large value, in such a way that the integral of this component taken across the gap defines the voltage. The component of **E** "round" the gap, on the other hand, may be supposed to vanish. This "model" is, effectively, that which has been adopted by Stratton and Chu for a perfectly conducting (spherical or spheroidal) antenna,[8] a detailed discussion of which has recently been given by Infeld.[9]

With an ordinary antenna in the form of a wire, there will be no difficulty in picturing this gap. To include the case where the antenna is of any shape, however, we shall define the gap more precisely as a portion of the antenna surface bounded by two closed curves which are close together. We take two orthogonal families of curves in the gap, the two curves bounding the gap being members of one family. At any point of the gap two directions are thus defined, which we may refer to as "across the gap" and "round the gap." Let these two directions be denoted by s, s' respectively, related in such a way that a right-handed rotation from s to s' takes us along the inward normal to the antenna surface. This picture of an antenna with a gap is illustrated on an exaggerated scale in Fig. 1.

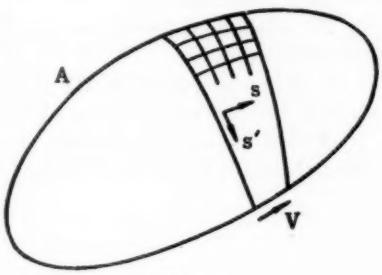


FIG. 1

By the "transmitting impedance" of an antenna, we shall mean, quite generally, the ratio of the current across the gap, or, more precisely, the current across ***one*** of the curves bounding the gap (and which we shall refer to simply as the "current at the gap"), to the impressed voltage (the positive sense of the current being the direction of voltage rise). This impedance will, in general, depend on what may be termed geometry of the gap (width, curvature, etc.)[10] and also on the details of the applied field, i.e. different applied fields, even with the same gap and the same voltage, may give different impedances. It is reasonable to assume, however, that, within certain limits as to the size of the gap,[11] the impedance is very nearly independent of the gap-geometry and of the details of the applied field, depending only on the position of the gap on the antenna surface. This is the situation which is usually tacitly assumed in engineering practice.

[7] It may be noticed that, in order to apply this concept in some cases, it may be necessary to extend the "antenna surface" beyond what is usually thought of as constituting the antenna. In the case of an antenna protruding through the wall of a wave-guide, for instance, the walls of the guide must (at least without further analysis) be treated as forming part of the antenna.

[8] J. A. Stratton and L. J. Chu, J. App. Phys., **12,** 241 (1941). These authors are only concerned with a transmitting antenna, but the same procedure can obviously be adopted for a receiving antenna.

[9] L. Infeld, Quart. Appl. Math. (appearing shortly).

[10] I am indebted to the referee for emphasizing to me that fact that impedance may depend on gap-width

[11] The gap must not be too small, for, as gap-width tends to zero, the investigations of Stratton and Chu (ref. 8) and Infeld (ref. 9) show that the reactive component of the current at the gap tends to infinity. On the other hand, the gap obviously must not be too large, or the details of the applied field become of importance.

It is not the purpose of this paper to investigate the conditions under which this "practical" case may be realized. The results to be proved, however, hold whether it is realized or not (the approximations made in Secs. 4 and 5 merely require that the width of the gap be not too great in order that they may be valid).

When the antenna is receiving, we have, similarly, an "induced voltage" and a "current at the gap." When the antenna is acting as a parasite, there is, of course, no gap. Since the same antenna may be imagined to act as a transmitter, parasite, or receiver, we shall, however, speak of the "current at the gap" even in the parasite case. By this we shall mean the current at what *becomes* the gap when the antenna acts as transmitter or receiver. Alternatively, we may regard the parasitic antenna as a special case of the receiving antenna when the gap is "filled up" (corresponding to zero load impedance).

We shall now formulate the three mathematical problems associated with an antenna as follows:

Transmission problem. We require a solution of Maxwell's equations in the region outside the antenna surface which satisfies the following conditions:

(1) On the antenna surface the tangential components of **E** vanish except in the gap, where the following relations must be satisfied:

$$E_{s'} = 0, \tag{1}$$

$$\int E_s ds = -V, \tag{2}$$

where the integral in (2) is taken across the gap and V is the applied voltage, which is supposed given (the integral is, of course, supposed to have the same value wherever it is taken across the gap).

(2) In the surrounding medium, the tangential components of **E** and **H** are continuous across any surfaces of discontinuity of the tensors $\epsilon_{ij}, \mu_{ij}, \sigma_{ij}$.

(3) At infinity only outgoing waves are present.[12]

Parasite problem. We are given an incident field (whose source is assumed to be unaffected by the presence of the antenna). We require a solution of Maxwell's equations for the *total* field (incident plus scattered) which satisfies the following conditions:

(1) On the antenna surface, the tangential components of **E** vanish *everywhere.*

(2) Same as condition (2) in transmission problem.

(3) At infinity, the field must reduce to the incident field together with only outgoing (or damped) waves.

Reception problem. We are again given an incident field. We require for the total field a solution of Maxwell's equations which satisfies the following conditions:

(1) On the antenna surface, the tangential components of **E** vanish except in the gap, where the following relations must be satisfied

$$E_{s'} = 0, \tag{3a}$$

$$\int E_s ds = Z_L I, \tag{3b}$$

[12] If the medium is conducting, the corresponding condition is that the field must be space-damped at infinity.

where I is the current at the gap and Z_L is the load impedance, i.e. the impedance at the gap looking *into* the connecting circuit, We regard Z_L as given.

(2), (3) Same as (2), (3) in parasite problem.

In accordance with what has already been said, the transmission problem, as formulated above, has not, strictly speaking, a unique solution. Instead of specifying the applied voltage V, in condition (1), we must specify the whole applied *field* E_s (a dependence of the solution on gap-geometry does not, of course, affect the *uniqueness* of solution). Again, in the reception problem, it is not sufficient to specify the load impedance Z_L. To make the problem determinate we must specify something as to the nature of the field E_s within the gap. But any such possible indeterminancy in the solution of the transmission or reception problems is immaterial for our purpose.

In the case of the parasite problem, there is no gap and no question as to the uniqueness of solution arises.

3. Solution of the reception problem in terms of the solutions of the transmission and parasite problems. Thévenin's theorem. It is now easy to show that the reception problem for any antenna can be solved very simply once the transmission and parasite problems for the same antenna have been solved, it being assumed that the gap is the same for the transmission and reception problems, and that the incident field is the same for the parasite and reception problems.

Let $(\mathbf{E}_1, \mathbf{H}_1)$ be the total field (incident plus scattered) in the parasite problem, and $(\mathbf{E}_2, \mathbf{H}_2)$ the field in the transmission problem when there is *unit* applied voltage (or one such field in the case where the transmission problem is not uniquely determined by specifying the voltage). Then a field which is a linear combination of these two fields of the form

$$\mathbf{E} = \mathbf{E}_1 + \alpha\mathbf{E}_2, \qquad \mathbf{H} = \mathbf{H}_1 + \alpha\mathbf{H}_2, \tag{4}$$

where α is a properly chosen constant, gives the solution of the reception problem (or *a* solution, if there is more than one).

For, in the first place, this field is a solution of Maxwell's equations, since all equations are supposed linear. Again, taking account of the conditions satisfied by the $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_2, \mathbf{H}_2)$ fields, it is evident that conditions (2) and (3) of the reception problem, and also condition (1) except in the gap, are satisfied. Further, in the gap we have, by virtue of the conditions satisfied there for the $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_2, \mathbf{H}_2)$ fields:

$$E_{s'} = 0, \tag{5}$$

$$\int E_s ds = \alpha \int E_{2,s} ds = -\alpha, \tag{6}$$

while the current at the gap for the $(\mathbf{E}, \mathbf{H})$ field is given by

$$I = I_1 + \alpha I_2 = I_1 + \alpha/Z_T, \tag{7}$$

where I_1, I_2 are the currents at the gap for the $(\mathbf{E}_1, \mathbf{H}_1)$, $(\mathbf{E}_2, \mathbf{H}_2)$ fields respectively, and Z_T is the transmitting impedance of the antenna as defined in the previous section (in the general case, we would have to say the transmitting impedance for the particular applied field chosen). From (5), (6) and (7) we see that condition (1) of the reception problem is satisfied in the gap if $\alpha = Z_L(I_1 + \alpha/Z_T)$, or

$$\alpha = -Z_L Z_T I_1/(Z_L + Z_T). \tag{8}$$

Hence (4) gives the solution of the reception problem in terms of the solutions of the transmission and parasite problems, since (8) determines α in terms of known quantities, Z_T being known from the solution of the transmission problem, I_1 (the current at the gap in the parasite problem) from the solution of the parasite problem, and Z_L being one of the data of the reception problem. We thus see that the solution of the reception problem as formulated in the previous section is unique if, with the given gap, the transmitting impedance Z_T is uniquely defined.

In particular, we find from (7) and (8) for the current at the gap in the reception problem

$$I = Z_T I_1/(Z_L + Z_T), \tag{9}$$

where I_1 is the current at the gap in the corresponding parasite problem. Hence the voltage which appears across the gap in the reception problem is

$$V = -Z_L I = -Z_L Z_T I_1/(Z_L + Z_T). \tag{10}$$

The results (9) and (10) show that Thévenin's theorem, suitably interpreted, holds for an antenna. It may be remarked that these results are rigorous with the assumptions here made and apply formally to an *arbitrary* gap provided that the impedance Z_T be understood in the general sense already explained.

4. Connection between the transmission and parasite problems. We have shown in the previous section that the reception problem may be solved very simply in terms of the solutions of the transmission and parasite problems. We shall now establish a connection between the transmission and parasite problems. To be precise, we shall show that *if the transmission problem has been solved, then the current at the gap in the parasite problem can be found, provided*: (1) *the incident field can be replaced by the field of an equivalent Hertzian (infinitesimal) dipole*, (2) *the medium, in the neighborhood of this dipole, is homogeneous and isotropic, being characterized by a dielectric constant* ϵ, *permeability* μ *and (possibly) conductivity* σ. These restrictions are practically always satisfied in practice. At some distance from the source of the incident field, this field is the same as that of an appropriately chosen Hertzian dipole (or a superposition of such dipoles) and it is only the incident field in the neighborhood of the parasitic antenna which is of significance for the satisfaction of the boundary conditions in the parasite problem. Again, the medium can always, at least to a sufficient approximation, be treated as homogeneous and isotropic in the neighborhood of the equivalent Hertzian dipole.

Thus we can determine the current at the gap in the parasite problem when once the transmission problem has been solved. The results (9) and (10) then give the current at the gap and the voltage across the gap in the reception problem (with the same incident field). By varying the gap in the transmission problem we can determine the current distribution everywhere in the antenna in the parasite problem.[13] Knowing this, it is usually possible to completely determine the field in the parasite problem (and hence also in the reception problem). Thus we may say that a complete solution of the transmission problem furnishes also a complete solution of the parasite and reception problems. However, in the reception problem, it is usually only the voltage induced across the gap which is of interest.

[13] This statement requires some modification in the case of an antenna of arbitrary shape, since what is actually furnished is the current across a closed curve on the antenna surface. For the usual case of an antenna in the shape of a wire, however, the statement remains true.

We shall now establish the result stated above, namely, that a solution of the transmission problem furnishes also the current at the gap in the parasite problem under the conditions stated. Our method of analysis is here similar to that of Sommerfeld,[3] except that we deal with a real antenna and a (fictitious) Hertzian dipole, whereas Sommerfeld considered two Hertzian dipoles. Our point of departure is the reciprocal theorem of Lorentz,[14] which may be stated as follows: If $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_2, \mathbf{H}_2)$ are two Maxwellian fields which have no singularities in the region external to any number of closed surfaces U, and which both give only outgoing (or damped) waves at infinity, then

$$\int_U (\mathbf{E}_1 \times \mathbf{H}_2)_n dS = \int_U (\mathbf{E}_2 \times \mathbf{H}_1)_n dS, \tag{11}$$

where n denotes (say) the outward normal to the surfaces U. The result (11) applies if the medium outside the surfaces U is of the general linear type specified at the beginning of Sec. 2. It was proved by Lorentz for the case of an isotropic medium, but the extension to the case of a crystalline medium is easily made; this is shown in the appendix.

Now let A denote the antenna surface (Fig. 1) and S a small sphere of radius r whose center is the point P where the equivalent Hertzian dipole is situated (Fig. 2).

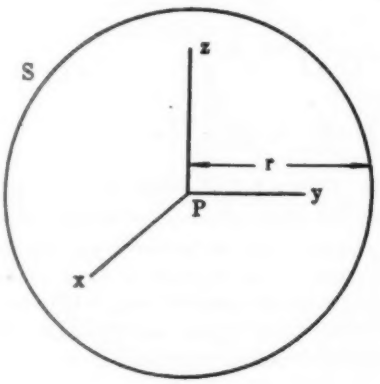


FIG. 2

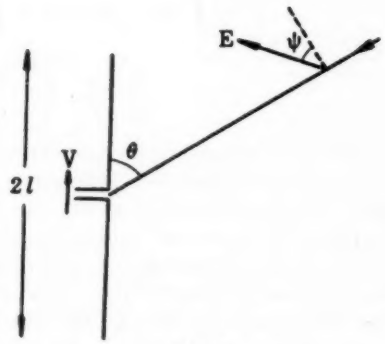


FIG. 3

Also, let $(\mathbf{E}_1, \mathbf{H}_1)$ denote the field (or *a* field) when the antenna is transmitting with *unit* current at the gap and no dipole is present at P, and $(\mathbf{E}_2, \mathbf{H}_2)$ the field when the dipole at P is transmitting and the antenna is acting as a parasite. Then, from (11), we have

$$\int_{A+S} (\mathbf{E}_1 \times \mathbf{H}_2)_n dS = \int_{A+S} (\mathbf{E}_2 \times \mathbf{H}_1)_n dS, \tag{12}$$

when n denotes the outward normal to A or S.

On using the boundary conditions satisfied by the $(\mathbf{E}_2, \mathbf{H}_2)$ field, we have at once

$$\int_A (\mathbf{E}_2 \times \mathbf{H}_1)_n dS = 0. \tag{13}$$

[14] H. A. Lorentz, Amsterdamer Akad. van Wetenschappen **4**, 176 (1895).

To calculate the other integral over A in (12), we introduce orthogonal curvilinear coordinates (ξ,η) in the gap, the curves $\eta=$constant being in the "s"-direction, and the curves $\xi=$constant in the "s'" direction (Fig. 1). Let the line element in the gap be

$$ds^2 = h_1(\xi, \eta)d\xi^2 + h_2(\xi, \eta)d\eta^2.$$

Then, taking account of the boundary conditions satisfied by the $(\mathbf{E}_1, \mathbf{H}_1)$ field, we have

$$\int_A (\mathbf{E}_1 \times \mathbf{H}_2)_n dS = \int\int E_{1,\xi} H_{2,\eta} h_1 h_2 d\xi d\eta,$$

where the double integral is taken over the surface of the gap, and the limits for ξ, η are constants.

We shall now assume that we can neglect the dependence of $H_{2,\eta}$ and h_2, or at least of the product $H_{2,\eta}h_2$, on ξ. This amounts to neglecting the variation of the current density across the gap when the antenna is acting as parasite, and to neglecting the width of the gap in comparison with the radius of curvature of the s-curves. The former involves an error which would usually be of the order w/λ, where w is the width of the gap and λ the wave-length. The latter involves an error of the order w/R where R is the radius of curvature of the s-curves.

With this assumption we now have

$$\begin{aligned}\int_A (\mathbf{E}_1 \times \mathbf{H}_2)_n dS &= \int E_{1,\xi} h_1 d\xi \cdot \int H_{2,\eta} h_2 d\eta \\ &= \int E_{1,s} ds \cdot \oint H_{2,s'} ds', \end{aligned} \tag{13a}$$

where the first integral in the last expression in (13a) is taken across the gap and the second integral is taken round the gap along any one of the s'-curves. The first integral is equal to the voltage across the gap when the antenna is transmitting, while the second integral (since the antenna is perfectly conducting) is equal to $4\pi/c$ times the current at the gap when the antenna is acting as a parasite (the difference between the currents across any one of the s'-curves is neglected, according to our assumption). Remembering that in the $(\mathbf{E}_1, \mathbf{H}_1)$ field, the applied voltage is assumed to give unit current at the gap, we thus finally have

$$\int_A (\mathbf{E}_1 \times \mathbf{H}_2)_n dS = \frac{4\pi}{c} Z_T I_1, \tag{14}$$

where Z_T is the transmitting impedance of the antenna, and I_1 is the current at the gap in the $(\mathbf{E}_2, \mathbf{H}_2)$ field.

We shall now calculate the integrals over S in (12) when we proceed to the limit $r\to 0$. For this purpose, let us take axes (Fig. 2) with the origin at P, and the z-axis in the direction of the axis of the dipole. Then, since the medium is supposed to be homogeneous and isotropic in the neighborhood of P, we have for the $(\mathbf{E}_2, \mathbf{H}_2)$ field on the sphere S, correct to terms of order $1/r^2$:

$$\left.\begin{aligned} &E_{2,x} = 3Mzx/r^5, \qquad E_{2,y} = 3Mzy/r^5, \qquad E_{2,z} = M\left(\frac{3z^2}{r^5} - \frac{1}{r^3}\right), \\ &H_{2,x} = ik\epsilon' My/r^3, \qquad H_{2,y} = -ik\epsilon' Mx/r^3, \quad H_{2,z} = 0, \end{aligned}\right\} \tag{15}$$

where

$$k = \omega/c, \qquad \epsilon' = \epsilon + 4\pi i\sigma/\omega, \tag{16}$$

and M denotes the strength[15] of the dipole. We thus see that $\mathbf{E}_2$ is of order $1/r^3$, whereas $\mathbf{H}_2$ is of order $1/r^2$, on S. Hence, to calculate the limit of the integrals over S occurring in (12), it is sufficient to put

$$\mathbf{E}_1 = (\mathbf{E}_1)_P, \tag{17}$$

but it is *not* sufficient to put $\mathbf{H}_1 = (\mathbf{H}_1)_P$. We must, instead, put

$$H_{1,x} = (H_{1,x})_P + x\left(\frac{\partial H_{1,x}}{\partial x}\right)_P + y\left(\frac{\partial H_{1,x}}{\partial y}\right)_P + z\left(\frac{\partial H_{1,x}}{\partial z}\right)_P, \tag{18}$$

with similar expressions for $H_{1,y}$, $H_{1,z}$.

Using (15), (17), (18), the calculation of the integrals over S is straightforward and similar to the corresponding calculations of Sommerfeld,[3] except that he omitted to take account of the variation of $\mathbf{H}_1$ over S (in our notation), as expressed by (18).[16] We thus find that, as $r \to 0$,

$$\int_S (\mathbf{E}_1 \times \mathbf{H}_2)_n dS = \frac{8}{3}\pi ik\epsilon' M(E_{1,z})_P, \tag{19}$$

$$\int_S (\mathbf{E}_2 \times \mathbf{H}_1)_n dS = \frac{4}{3}\pi M\left(\frac{\partial H_{1,y}}{\partial x} - \frac{\partial H_{1,x}}{\partial y}\right)_P = -\frac{4}{3}\pi ik\epsilon' M(E_{1,z})_P. \tag{20}$$

The last step in (20) follows from the Maxwell equation

$$\text{curl } \mathbf{H}_1 = -ik\epsilon'\mathbf{E}_1,$$

which, by supposition, holds in the neighborhood of P.

Substituting (13), (14), (19), (20) in (12), we now have

$$Z_T I_1 = -i\omega\epsilon' M E_d, \tag{21}$$

where $E_d = (E_{1,z})_P$ denoted the component along the axis of the equivalent dipole of the electric field generated when the antenna is transmitting with unit current at the gap. We repeat here that M is the strength of the equivalent Hertzian dipole, and ϵ' the "complex dielectric constant," defined by (16), of the medium at the position of the equivalent dipole. This is the required result. For, if the transmission problem has been solved, Z_T and E_d are known, and (21) then gives I_1, the current at the gap in the parasite problem. In general, both Z_T and E_d might depend on the details of the applied field, but (21) shows that, to the approximation with which (13a) holds, the ratio Z_T/E_d is independent of such details.[17]

The position and strength of the equivalent dipole can usually be assigned very

[15] The "strength" of the dipole can be defined by saying that the field is given by (15) in the immediate neighborhood of the dipole; alternatively, by saying that the field is that derived from a Hertz vector whose components are $(0, 0, M\exp(ik\sqrt{\epsilon'\mu}r)/r$.

[16] Sommerfeld thus obtained the incorrect result zero for the integral in (20). This error does not appear in his final result, since he was dealing with two Hertzian dipoles. It is essential to correct it in the present calculations, however.

[17] The case of unit current at the gap has been chosen simply for convenience. In general, the ratio V/E_d' appears in place of the ratio Z_T/E_d in (21), where V is the applied voltage and E_d' the corresponding value of E_d.

easily when once the incident field is given. The assignment is usually by no means unique. In the common case of a plane-polarized wave, the equivalent dipole can conveniently be taken at a large distance from the antenna, and the strength of the dipole expressed in terms of the amplitude of the incident wave. If the incident field is the same as that of *two* equivalent Hertzian dipoles which are out of phase, as is the case with an elliptically-polarized wave, we need merely superpose the fields of each dipole separately, obtaining in place of (21)

$$Z_T I_1 = - i\omega\epsilon'(M_1 E_{d1} + M_2 E_{d2}), \tag{22}$$

where the suffixes 1 and 2 on the right-hand side refer to the two equivalent dipoles.

Let us, for instance, apply (21) to the case where the *whole* medium is homogeneous and isotropic, being characterized by constants ϵ, μ, σ, and where the incident field is that of a plane-polarized wave. Let us take polar coordinates (R, θ, ϕ) with the origin in the neighborhood of the antenna, and let us take the equivalent dipole at a large distance from the antenna in the direction from which the incident wave arrives, the axis of the dipole being in the direction of the electric vector of the incident wave. Then if E_{inc} is the (complex) amplitude of the electric vector in the incident wave at the origin, we have

$$E_{\text{inc}} = k'^2 M e^{ik'R}/R, \tag{23}$$

where ϵ' is again the "complex dielectric constants" defined by (16), and $k' = \sqrt{\epsilon'\mu}k$. Again, for *any* shape of antenna, we can always express E_d in the form

$$E_d = i\mu f(\theta, \phi, \psi) e^{ik'R}/cR, \tag{24}$$

where ψ is an azimuthal angle with respect to the radius vector which defines the direction of the electric vector in the incident wave, and $f(\theta, \phi, \psi)$ is a dimensionless function which can be calculated when the transmission problem of the antenna has been solved. This function determines the directional and polarization characteristics of the antenna when transmitting. Substituting for M and E_d from (23), (24) in (21), we have for the current at the gap in the parasite problem in the present case

$$I_1 = E_{\text{inc}} f(\theta, \phi, \psi)/kZ_T. \tag{25}$$

Equation (25) gives the current at the gap in the parasite problem in terms of the amplitude of the incident wave and quantities which are known from the solution of the transmission problem. From (10), we then have for the voltage induced at the gap in the reception problem

$$V = - Z_L E_{\text{inc}} f(\theta, \phi, \psi)/k(Z_L + Z_T). \tag{26}$$

Consider, for example, the well-known case of a linear antenna of length $2l$, with the gap at the middle point. Let us take ψ to be the angle between the electric vector of the incident wave and the plane containing the antenna and the direction of the incident wave (Fig. 3). Also let us take the origin at the center of the antenna with the polar axis along the antenna (in the direction in which the voltage rise is measured), so that θ is the angle which the direction of the incident wave makes with the antenna. Then, with the usual simplified treatment and the assumption of a sinusoidal current in the antenna for the transmission problem, we have for the function f in (24)

$$f(\theta, \phi, \psi) = 2 \frac{\cos (kl \cos \theta) - \cos kl}{\sin kl \sin \theta} \cos \psi. \tag{27}$$

Hence (26) now gives for the voltage induced at the gap in the reception problem

$$V = -\frac{2}{k} \frac{Z_L E_{\text{inc}}}{Z_L + Z_T} \frac{\cos (kl \cos \theta) - \cos kl}{\sin kl \sin \theta} \cos \psi. \tag{28}$$

The result (28) is in agreement with one obtained by Hallén[18] by a direct treatment of the reception problem, to the order to which (27) is true, i.e., with the usual assumption of a sinusoidal current in the antenna when transmitting. A more accurate treatment of the transmission problem would yield a more accurate expression than (27) for the function f, and hence also a more accurate expression than (28) for the induced voltage in the reception problem.

5. Reciprocal relations between two antennas. The results of the two preceding sections establish relations between the performance of a single antenna when acting as transmitter, parasite or receiver. They deal with the *fields* generated or received by the antenna, and make no mention of the second antenna which is required to detect the transmitted field, or to generate the incident field. Such relations would seem to be those of most use in establishing a connection between the transmitting and receiving properties of antennas. The more usual results establishing this connection, however, such as have been given by Carson[1] and Sommerfeld,[3] relate to the performance of *two* antennas. We shall now establish such "reciprocal relations."

Consider two antennas, A_1, A_2, with gaps G_1, G_2 (Fig. 4). Let us apply the Lorentz

FIG. 4

relation (11) to the region outside A_1, A_2, taking for $(\mathbf{E}_1, \mathbf{H}_1)$ the field (or a field) when A_1 is transmitting with an applied voltage V_1 and A_2 acting as parasite, and for $(\mathbf{E}_2, \mathbf{H}_2)$ the field when A_2 is transmitting with an applied voltage V_2 and A_1 acting as parasite. By an exactly similar method to that used in deriving (13a) and (14), and with a similar approximation (i.e. neglecting terms of order w/λ and w/R) we now have:

$$\int_{A_1} (\mathbf{E}_1 \times \mathbf{H}_2)_n dS = \frac{4\pi}{c} I_1 V_1, \tag{29}$$

$$\int_{A_2} (\mathbf{E}_2 \times \mathbf{H}_1)_n dS = \frac{4\pi}{c} I_2 V_2, \tag{30}$$

$$\int_{A_2} (\mathbf{E}_1 \times \mathbf{H}_2)_n dS = \int_{A_1} (\mathbf{E}_2 \times \mathbf{H}_1)_n dS = 0, \tag{31}$$

[18] E. Hallén, Nova Acta Reg. Soc. Upsaliensis **11**, No. 5 (1938).

where I_1 is the current induced at G_1 in the $(\mathbf{E}_2, \mathbf{H}_2)$ field, and I_2 the current induced at G_2 in the $(\mathbf{E}_1, \mathbf{H}_1)$ field. Substituting (29), (30), (31), in (11) we obtain[19]

$$V_1 I_1 = V_2 I_2. \tag{32}$$

This is the required reciprocal relation. We may state the content of the relation in words as follows: Let A_1, A_2 *be two antennas with gaps* G_1, G_2 *respectively. Then the current induced at* G_1 *when* A_2 *is transmitting and* A_1 *acting as parasite is equal to the current induced at* G_2 *when* A_1 *is transmitting and* A_2 *acting as parasite, providing that the applied voltages in the two cases are equal.*[20] This is the more precise form of Carson's[1] original reciprocal theorem.

More generally, if each antenna acts as a *receiver* instead of a parasite when the other is transmitting, let Z_{T1} be the transmitting impedance of A_1, and Z_{L1} the load impedance when A_1 is receiving, the gap being at G_1 in both cases. Let Z_{T2}, Z_{L2} denote similar quantities for A_2, the gap being at G_2. Then (9) and (32) lead to the relation

$$(1 + Z_{L1}/Z_{T1})V_1 I_1 = (1 + Z_{L2}/Z_{T2})V_2 I_2 \tag{33}$$

in place of (32), where I_1 is now the current induced at G_1 when A_2 is transmitting with applied voltage V_2 and A_1 receiving, and I_2 is the current induced at G_2 when A_1 is transmitting with applied voltage V_1 and A_2 is receiving. If, instead of induced currents, we deal with induced voltages, the relation (33) becomes

$$(Z_{T1} + Z_{L1})V_1 v_1 = (Z_{T2} + Z_{L2})V_2 v_2, \tag{34}$$

where v_1 is the voltage induced at G_1 when A_1 is receiving and A_2 transmitting, and v_2 is the voltage induced at G_2 when the roles of transmitter and receiver are interchanged. V_1, V_2 denote, as before, the *applied* voltages.

We may here briefly refer to an apparent contradiction of the relation (32) which has been put forward by Ballantine.[2] It was originally given by this author in criticism of Sommerfeld's[3] form of reciprocal theorem; but it would apply equally well to the more exact form of the theorem given above. Let the antenna A_1 considered in the relation (32) be a linear whole-wave antenna which, when transmitting, has a sinusoidal current distribution which is antisymmetrical about the middle point ("complete positive and negative current loops"), as shown in Fig. 5. Let the second antenna A_2, which may be of arbitrary shape,[21] be situated at a large distance from A_1 and in a direction perpendicular to it (Fig. 5). Then, when A_1 is transmitting, with the assumed current distribution, the field-strength generated is zero in the direction of A_2,[22] so that no current will be induced anywhere in A_2 when A_2 is acting as parasite. On the other hand when A_2 is transmitting and A_1 acting as parasite it is certainly

[19] In the notation more generally used, I_1 and I_2 are interchanged in (32).

[20] It should be recalled that we are dealing with complex amplitudes. What is implied, therefore, is that if the (real) amplitudes of the applied voltages are equal in the two cases, then both the current amplitudes and the phase differences between induced current and applied voltage are also equal. The frequencies must, of course, be the same in the two cases.

[21] Ballantine supposes that the second antenna is exactly similar to the first (but not parallel to it) in order to comply with the conditions of Sommerfeld's theorem. But this restriction is of no importance in the present connection.

[22] The longitudinal component of field-strength is actually only zero when terms of order higher than $1/r$, where r denotes distance from the antenna, are neglected. This is all that is required for the argument, however.

not true in general that no current is induced in A_1. This contradicts the reciprocal relation (32).

The explanation is that a completely antisymmetrical current distribution of the type postulated when A_1 is transmitting cannot occur. Ballantine makes no mention as to where the antenna is supposed to be fed. If fed at the center, the current distribution is *symmetrical*, not antisymmetrical. If fed at some other point, well away from the center, the current distribution will be *approximately* of the type postulated, but not exactly, since the usual sinusoidal theory is not exact. The field-

FIG. 5

strength generated by A_1 in the direction of A_2 is thus *not* zero, but is comparatively small. The reciprocal relation then tells us (quite correctly) that A_1 will be comparatively weakly affected when A_2 is transmitting.

6. Are the directional properties (polar diagrams) of an antenna identical for transmission and reception? It is commonly stated that the directional properties of an antenna, as exhibited in a polar diagram, are identical for transmission and reception. It is of interest to see whether the results of the two preceding sections confirm this or not.

The reciprocal relation (32) shows at once that, in a certain sense, the directional properties are identical. Consider an antenna A and a large number of other antennas (which need not necessarily be identical) situated at a large distance from A in different directions. Then we might define the polar diagram of A for transmission by means of the currents induced at certain "gaps" in these antennas when they act as parasites, and the polar diagram of A for reception by means of the currents induced at the gap in A when A acts as parasite and any one of the distant antennas is transmitting, the same voltage being applied in every case. The relation (32) then shows that the polar diagrams are the same.

Again, instead of considering the antennas as acting as parasites, we might consider them as acting as *receivers*, the load impedance being the same for all the distant antennas, and use the induced voltages, rather than the parasitic currents, to define the polar diagrams for transmission and reception. The relation (34) then shows that the polar diagrams are identical (v_1/v_2 constant for V_1/V_2 constant) only if the transmission impedances Z_T are identical for all the distant antennas. If all these antennas are identical, this would only be the case if the whole medium is homogeneous and isotropic; otherwise, the equality of the transmission impedances could be achieved by proper choice of the distant antennas.

Neither of the above ways of defining the polar diagrams is very satisfactory, however, and neither of them corresponds to what is usually adopted in practice. The polar diagram for transmission is usually defined by means of the field generated, or

the intensity, at distant points, and the polar diagram for reception by means of the voltage induced by, or the power absorbed from, an incident wave of given strength arriving from different directions.

For definiteness we shall, in an endeavor to treat the matter quite generally, define the polar diagrams of an antenna for transmission and reception (with the same gap) as follows: draw a large sphere S[23] surrounding the antenna A. Then the polar diagram for transmission is constructed by making the length of the radius vector in any direction in the diagram proportional to the maximum amplitude of the electric field-strength generated (for constant applied voltage[23a]) at the point on S which lies in this direction. For reception, we suppose that the source of the incident field can be replaced by an equivalent Hertzian dipole (cf. Sec. 4) at some (variable) point P on S, the axis of the dipole being in the direction in which the electric field-strength generated at P when A is *transmitting* is a maximum.[24] The strength of the equivalent dipole is then adjusted so that the maximum amplitude of the electric field-strength in the incident field at A is the same for all points P on S. Then the polar diagram for reception is constructed by making the length of the radius vector in the direction of P in the diagram proportional to the voltage induced in A when the equivalent dipole is at P, the gap in A being the same as for transmission and the load impedance being kept constant. It is perhaps worth pointing out that the direction of P (i.e., of the source) is not necessarily the same as the direction of arrival of the incident wave if inhomogeneities of the medium are present. The incident field at the antenna need not, in fact, be a plane wave at all, in which case we cannot even speak of the direction of the incident wave.

With these definitions, it follows from (21) and (10) that the polar diagrams for transmission and reception are identical if, and only if, the expression $|\epsilon' M/E_{\text{inc}}|$ is constant for all directions from the antenna (i.e. for all points on the distant sphere), where $|E_{\text{inc}}|$ denotes the maximum amplitude of the electric field-strength in the incident field when the strength of the equivalent Hertzian dipole is M, and ϵ' is the complex dielectric constant of the medium at the position of this dipole.

The expression $|\epsilon' M/E_{\text{inc}}|$ is certainly constant if the whole medium is homogeneous and isotropic, for ϵ' and M/E_{inc} are then both constant. It is also constant, for instance, if we are considering transmission and reception over a perfectly conducting earth, or over an imperfectly conducting, but homogeneous, earth. It is *not* in general constant, however, when inhomogeneities are present—e.g. for a heterogeneous imperfectly conducting earth.

We conclude, therefore—at least with the above definitions of the polar diagrams—that the identity of the polar diagrams is not a *general* property of antennas when inhomogeneities of the medium are present, although the diagrams are probably identical, or approximately so, in practically-occurring cases.

We may note incidentally that (21) and (10) show that the voltage induced in an antenna by an equivalent Hertzian dipole of given moment, and situated at a given point, is a maximum when the axis of the dipole lies in the direction in which the electric field-strength generated at that point when the antenna is transmitting is a

[23] In general, the polar diagrams obtained from our definitions may depend on the radius of this sphere. In practice they will usually not do so, however, provided that the sphere is *large*.

[23a] The diagram thus obtained is independent of the applied voltage by (21) and Footnote 17.

[24] This somewhat arbitrary choice is made so as to render conditions most favorable for the identity of the two polar diagrams. It implies, of course, a certain restriction on the source of the incident wave.

maximum. When the whole medium is homogeneous and isotropic, this means that the voltage induced by a plane-polarized incident wave is a maximum, other things being equal, when the polarization of the incident wave is the same as that of the wave generated when the antenna is transmitting (or, in the general case, where the generated wave is elliptically-polarized, when the electric vector of the incident wave is in the direction of the major axis of the ellipse described by the electric vector of the generated wave).

7. Extension of the results to include imperfect conductivity of the antenna. So far, we have dealt entirely with the case where the antenna is perfectly conducting. We shall now indicate briefly how the results can be extended to include the case of an imperfectly conducting antenna.

We shall now adopt the following "mathematical model" for a transmitting or receiving antenna (cf. Sec. 2): there is a "gap" as before, and the conditions (1), (2) or (3a), (3b) again hold; but we now postulate in addition that over the *ends* of the gap (i.e., the cross-sections of the antenna through the curves marking the position of the gap) the tangential components of **E** vanish, both in transmission and reception. The "gap" is thus now a kind of "box," and the surrounding medium now includes the interior of the antenna which lies outside this "box." The transmission and reception problems for an antenna now require a solution of Maxwell's equations in the region outside the "box," subject to the proper boundary conditions in this region, which are the same as before, and subject to the appropriate surface boundary conditions on the box. The latter are the same as before, with the additional condition that the tangential components of **E** vanish over the ends of the box. The condition that the tangential components of **E** vanish on the surface of the antenna now disappears. In the parasite problem, there is no "box" and hence, of course, no conditions to be applied to its surface.

We shall now make the followed additional assumptions:

(1) The displacement current at the gap is negligible compared with the conduction current.

(2) When the antenna is acting as a parasite, the integral $\int E_s ds$, taken across the gap (Fig. 1), is negligible.

(3) When the antenna is acting as a parasite, the component $E_{s'}$ (Fig. 1) and the tangential components of **E** over the ends of the gap are negligible.

The only one of these assumptions to which exception might be taken is the last, which is certainly *not* true for an antenna of arbitrary shape and with an arbitrary gap. It is, however, very approximately true for the usual wire-shaped antenna, so that it seems reasonable to adopt it as a general assumption, particularly in view of the somewhat arbitrary character of our "box model."

It is now readily verified that the general results of sections 3–5 can be proved as before, by merely substituting the surface of the box for the surface of the antenna (not all the assumptions (1)–(3) above are needed in the proof of each result). The discussion of section 6 then applies as before. The formula (27) (and hence (28)), which relates to a special problem, requires modification, of course, if the imperfect conductivity of the antenna is taken into account.

Appendix

We shall here establish the Lorentz relation (11) when the medium outside the surfaces U is anisotropic, being characterized by the symmetric tensors ϵ_{ij}, μ_{ij}, σ_{ij},

which may be arbitrary functions of position (cf. Sec. 2). Adopting for a moment the indicial notation, so that, e.g., we write (E_1, E_2, E_3) in place of (E_x, E_y, E_z), Maxwell's equations for the steady state (with the time-factor $e^{-i\omega t}$), and for the postulated type of medium, can be written

$$(\text{curl } \mathbf{E})_i = -\frac{i\omega}{c} \sum_j \mu_{ij} H_j,$$

$$(\text{curl } \mathbf{H})_i = \frac{i\omega}{c} \sum_j \epsilon_{ij} E_j + \frac{4\pi}{c} \sum_j \sigma_{ij} E_j,$$

for $i = 1, 2, 3$, the summations running in all cases from 1 to 3.

Suppose now we have two fields $(\mathbf{E}^{(1)}, \mathbf{H}^{(1)})$ and $(\mathbf{E}^{(2)}, \mathbf{H}^{(2)})$ which satisfy these equations. Then we can write:

$$(\text{curl } \mathbf{E}^{(1)})_i = \sum_j \alpha_{ij} H_j^{(1)}, \tag{A1}$$

$$(\text{curl } \mathbf{H}^{(1)})_i = \sum_j \beta_{ij} E_j^{(1)}, \tag{A2}$$

$$(\text{curl } \mathbf{E}^{(2)})_i = \sum_j \alpha_{ij} H_j^{(2)}, \tag{A3}$$

$$(\text{curl } \mathbf{H}^{(2)})_i = \sum_j \beta_{ij} E_j^{(2)}, \tag{A4}$$

where

$$\alpha_{ij} = -\frac{i\omega}{c} \mu_{ij}, \qquad \beta_{ij} = \frac{i\omega}{c} \epsilon_{ij} + \frac{4\pi}{c} \sigma_{ij}. \tag{A5}$$

Multiplying (A1), (A2), (A3, (A4) by $H_i^{(2)}$, $E_i^{(2)}$, $-H_i^{(1)}$, $-E_i^{(1)}$ respectively, adding, and summing with respect to i, we obtain, on using the identity $\mathbf{A} \cdot \text{curl } \mathbf{B} - \mathbf{B} \cdot \text{curl } \mathbf{A} = \text{div } (\mathbf{B} \times \mathbf{A})$:

$$\text{div } [\mathbf{E}^{(1)} \times \mathbf{H}^{(2)} - \mathbf{E}^{(2)} \times \mathbf{H}^{(1)}]$$

$$= \sum_{i,j} [\alpha_{ij}(H_j^{(1)} H_i^{(2)} - H_j^{(2)} H_i^{(1)}) + \beta_{ij}(E_j^{(1)} E_i^{(2)} - E_j^{(2)} E_i^{(1)})]$$

$$= \sum_{i,j} [(\alpha_{ij} - \alpha_{ji}) H_j^{(1)} H_i^{(2)} + (\beta_{ij} - \beta_{ji}) E_j^{(1)} E_i^{(2)}] = 0,$$

since, from (A5), the tensors α_{ij}, β_{ij} are symmetric.

Hence, using the notation of (11),

$$\text{div } (\mathbf{E}_1 \times \mathbf{H}_2) = \text{div } (\mathbf{E}_2 \times \mathbf{H}_1). \tag{A6}$$

From (A6) the required relation (11) follows in the usual way by integrating each side of the equation over the region bounded by the surfaces U and an infinite sphere, and applying Green's theorem. This sphere, and also any surfaces of discontinuity, give no contributions to the surface integrals on account of the boundary conditions (outgoing waves at infinity; continuity of tangential components of $\mathbf{E}$ and $\mathbf{H}$ at a surface of discontinuity).

DISTRIBUTION OF THE SUM OF RANDOMLY PHASED COMPONENTS

BY

W. R. BENNETT

Bell Telephone Laboratories, Murray Hill, N. J.

The so-called problem of random flights has engaged the attention of many investigators. In the general form we are required to find the distribution of the sum of n vectors with specified lengths and randomly distributed directional angles. From the standard application of complex algebra to sinusoidal waves, it is clear that the two-dimensional form of the problem, with which we shall be concerned here, applies directly to the distribution of the sum of n sine waves with equal frequencies, specified amplitudes, and random phase angles. By a simple extension which is reproduced in the appendix, we may also show equivalence with the distribution of the envelope of n sine waves of incommensurable frequencies when all frequencies present are high relative to the difference between the lowest and highest. Applications thus exist to such diverse problems as random walks, in terms of which Pearson originally phrased the problem in 1905,[1] and multichannel carrier telephony, which happens to be the stimulating influence behind the present investigation.

A concise summary of the general problem and its solution has been given by Watson.[2] From these results, we take Kluyver's formula for the distribution, namely

$$P_n(r) = r\int_0^\infty J_1(rt)dt\prod_{m=1}^{n} J_0(a_m t) \tag{1}$$

where $P_n(r)$ is the probability that the sum of n components having individual magnitudes a_1, $a_2 \cdots a_n$ is less than r. We point out that the special case of $n=1$ is trivial, the case of $n=2$ is resolvable in terms of elementary functions, and the case of $n=3$ can be evaluated by elliptic integrals,[3] so that the infinite integral presents no real computing problem for n less than four. For large values of n, the integral may be calculated satisfactorily from asymptotic series in the manner indicated by Rayleigh.[4] There remains considerable difficulty, however, in obtaining accurate numerical results for intermediate values of n. That the asymptotic series becomes not only unwieldy but unreliable in such cases is indicated by demonstrable errors in the results obtained by its use in one published paper.[5] It is difficult to obtain accuracy from mechanical or numerical integration because of the oscillating integrand.

The author recently wanted numerical values of $P_n(r)$ for values of n of the order of 10. Since the accuracy of the asymptotic series and of mechanical or numerical integration was open to suspicion, a better method of evaluating the integral (1) was

[1] K. Pearson, Nature **72,** 294 and 342 (1905).

[2] G. N. Watson, *Theory of Bessel functions*, Cambridge, Univ. Press, 1922, pp. 419–421.

[3] Nicholson, *Quart. J. Pure Appl. Math.*, **48,** 321–329 (1920).

[4] Phil. Mag. (6) **37,** 321–347 (1919).

[5] G. D. Gillett, *Some developments in common frequency broadcasting*, Proc. I.R.E. **19,** 1347–1369 (1931). Certain of the curves of Fig. 9 of this article fail to satisfy the easily proved relation $P_n(a)=1/(n+1)$ when $a_1=a_2=\cdots=a_n=a$.

sought. It was found possible to derive a convergent Fourier-Bessel series representation, which is simple in form and yields any accuracy desired. Derivation of the convergent series equivalent introduces some rather interesting analytical procedures and leads to a general method of evaluating certain types of Bessel integrals, particularly those which can be shown to vanish for all values of a parameter outside a certain finite range.

In the interests of simplicity we shall describe the method in terms of the special case of all vectors equal in length. The general case can be treated in a similar manner with more cumbersome details, of which only the final result will be stated here. We shall find it more convenient to calculate the probability Φ_n that the sum exceeds r. We set all the quantities $a_m (m=1, 2, \cdots n)$ equal to each other and without loss of generality make them each equal to unity. Then

$$\Phi_n(k) = 1 - P_n = 1 - k\int_0^\infty J_1(kt)J_0^n(t)dt = n\int_0^\infty J_0(kt)J_1(t)J_0^{n-1}(t)dt. \tag{2}$$

The last form follows from an integration by parts. $\Phi_n(k)$ is the probability that the sum of n vectors of equal lengths exceeds k times the length of one vector. The equation (2) holds for all real non-negative values of k. It follows that $\Phi_n(k)$ vanishes for $k>n$ since the sum of n equal vectors cannot be greater than n times the length of one. The vanishing of $\Phi_n(k)$ for $k>n$ can also be shown directly from the infinite integral.

Let us multiply both sides of (2) by $kJ_0(xk)dk$ and integrate with respect to k from zero to infinity. We have the formal result:

$$\int_0^\infty kJ_0(xk)\Phi_n(k)dk = \int_0^\infty kJ_0(xk)dk\int_0^\infty tJ_0(kt)F(t)dt \tag{3}$$

where

$$F(t) = nJ_1(t)J_0^{n-1}(t)/t. \tag{4}$$

The integral on the left certainly exists and in fact satisfies sufficient conditions as a function of x for the Fourier-Bessel integral identity to hold.[6] It follows that:

$$\int_0^\infty kJ_0(xk)\Phi_n(k)dk = F(x) = \frac{nJ_1(x)}{x}J_0^{n-1}(x). \tag{5}$$

We have thus found the Fourier-Bessel transform of $\Phi_n(k)$, or the Bessel analogue of the ordinary Fourier transform. We note that since Φ_n vanishes when k exceeds n, we may also write:

$$\int_0^n kJ_0(xk)\Phi_n(k)dk = \frac{nJ_1(x)}{x}J_0^{n-1}(x) \tag{6}$$

or if we let $k=nz$,

$$\int_0^1 zJ_0(nxz)\Phi_n(nz)dz = \frac{J_1(x)}{nx}J_0^{n-1}(x). \tag{7}$$

---

[6] G. N. Watson, *loc. cit.*, pp. 453–456.

We now let $\Phi_n(nz)=f(z)$ and note that $f(z)$ in the interval $0\leqq z\leqq 1$ satisfies a sufficient set of conditions[7] for its representation as the sum of a convergent Fourier-Bessel series:

$$f(z) = \sum_{m=1}^{\infty} c_m J_0(j_m z), \qquad 0 \leqq z \leqq 1 \tag{8}$$

where j_m is the mth root, in ascending order of magnitude, of the equation $J_0(x)=0$. The value of the coefficient c_m is given by:

$$c_m = \frac{2}{J_1^2(j_m)} \int_0^1 z f(z) J_0(j_m z)\, dz = \frac{2J_1(j_m/n)J_0^{n-1}(j_m/n)}{j_m J_1^2(j_m)}. \tag{9}$$

The last form of the result is obtained by substituting the value $x=j_m/n$ in (7). Hence

$$\Phi_n(k) = f(k/n) = 2\sum_{m=1}^{\infty} \frac{J_1(j_m/n)J_0^{n-1}(j_m/n)}{j_m J_1^2(j_m)} J_0(kj_m/n). \tag{10}$$

This is the result sought. It will be noted that $\Phi_n(n)=0$, as it should, since each term of the series then contains the factor $J_0(j_m)=0$. Incidentally, the vanishing of $\Phi_n(k)$ at $n=k$ is difficult if not impossible to show by asymptotic expansion. Values of $\Phi_n(k)$ computed from (10) are listed in Table I and plotted in Fig. 1.

TABLE I

Value of $\Phi_n(ka) = 1 - k\int_0^a J_1(k\,)J_0^n(z)dz = n\int_0^\infty J_0(kz)J_1(z)J^{n-1}(z)dz$

=probability that envelope of n sine waves of equal amplitude exceeds k times amplitude of one:

k	Φ_6	Φ_7	Φ_8	Φ_9	Φ_{10}	Φ_{100}
1	.8571	.8750	.8889	.9000	.9091	
2	.5335	.5818	.6211	.6537	.6812	
3	.2315	.2866	.3355	.3788	.4171	
4	.0636	.0992	.1356	.1712	.2055	.8528
5	.0088	.0515	.0384	.0581	.0793	
5.8	.00013					
6	—	.0021	.0067	.0137	.0229	
6.6		.00015				
7		—	.00050	.0020	.0046	.6138
7.2			.00025			
7.5			.00007			
8				.00012	.00056	
9				—	.00005	
10					—	.3688
15						.1052
20						.01795
25						.001802
30						.000104

[7] G. N. Watson, *loc. cit.*, pp. 591–592.

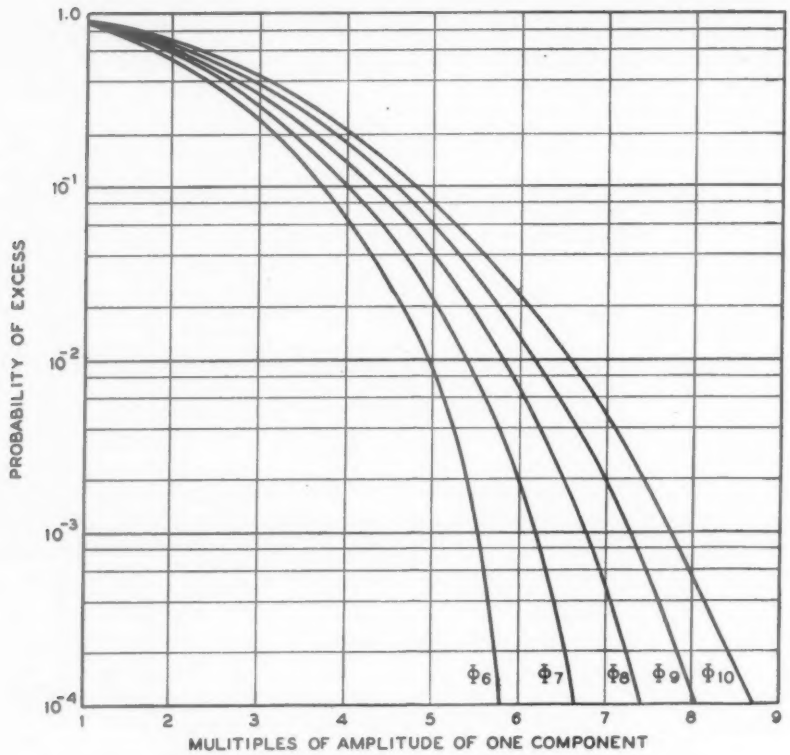


FIG. 1. $\Phi_n(k)$ = probability that the absolute value of the sum of n vectors of equal magnitude and random phase exceeds k times the absolute value of one component.

If the lengths of the vectors are unequal, the solution may be effected in a similar manner with the following results: Let

$$A = \sum_{m=1}^{n} a_m, \tag{11}$$

where a_m is the length of the mth vector. Also let

$$b_m = \sum_{s=1}^{n} J_0(a_s j_m / A), \tag{12}$$

$$c_m = \sum_{s=1}^{n} \frac{a_s J_1(a_s j_m / A)}{J_0(a_s j_m / A)}. \tag{13}$$

Then the probability that sum equals or exceeds r is:

$$\Phi_n = \frac{2}{A} \sum_{m=1}^{\infty} \frac{b_m c_m J_0(r j_m / A)}{j_m J^2(j_m)}. \tag{14}$$

A general mathematical theorem of which the above results are special cases may be stated as follows:

If

$$\int_0^\infty J_\nu(rt)F(t)t\,dt = 0 \quad \text{for} \quad r > r_0 \tag{15}$$

and if $F(t)$ satisfies conditions enabling its representation by the Fourier-Bessel integral; then for $0<r<r_0$,

$$\int_0^\infty J_\nu(rt)F(t)t\,dt = \frac{2}{r_0^2}\sum_{m=1}^{\infty}\frac{F(j_m/r_0)J_\nu(j_m r/r_0)}{J_{\nu+1}^2(j_m)} \tag{16}$$

where

$$J_\nu(j_m) = 0, \qquad j_1 < j_2 < j_3 < \cdots . \tag{17}$$

This completes the portion of our results dealing with the problem of random flights per se. It so happens that the primary object of the author is the application of the results to the envelope of a group of sine waves. As shown in the appendix the distribution function for the envelope of n randomly phased sine waves of independent frequencies distributed over a range narrow compared with the frequency of any component is in fact the same Φ_n calculated above. It is of interest to compare the distribution of the envelope of a group of sine waves with the distribution of the instantaneous values. The latter distribution can be calculated by the method of characteristic functions.[8] The probability density of the instantaneous value E of the sine wave $a_m \cos \omega t$ is given by[9]

$$p(E) = (a_m^2 - E^2)^{-1/2}/\pi, \qquad |E| < E_0. \tag{18}$$

The "characteristic function" (or Fourier transform) of $p(E)$ is

$$S_m(i\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} p(E)e^{-i\omega E}dE = \frac{J_0(a_m\omega)}{2\pi}. \tag{19}$$

For n sine waves with peak amplitudes $a_1, a_2, \cdots, a_n$, the probability density is given by

$$p_n(E) = (2\pi)^{n-1}\int_{-\infty}^{\infty} e^{iE\omega}d\omega\prod_{m=1}^{n} S_m(i\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{iE\omega}d\omega\prod_{m=1}^{n} J_0(a_m\omega). \tag{20}$$

This integral is not in convenient form to use the Fourier-Bessel integral and series. However, the presence of the exponential factor enables us to use the ordinary Fourier integral and series as follows: Multiply both sides of (20) by $e^{-itE}dE$ and integrate from $-\infty$ to ∞. Then

$$\int_{-\infty}^{\infty} p_n(E)e^{-itE}dE = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itE}dE\int_{-\infty}^{\infty} e^{iE\omega}d\omega\prod_{m=1}^{n} J_0(a_m\omega) = \prod_{m=1}^{n} J_0(a_m t) \tag{21}$$

by the Fourier integral identity. Since $p_n(E) = 0$ for $|E| > A$,

$$\int_{-A}^{A} p_n(E)e^{-itE}dE = \prod_{m=1}^{n} J_0(a_m t). \tag{22}$$

[8] H. Cramér, *Random variables and probability distributions*, Cambridge Tract No. 36 (1937), Ch. IV.
[9] W. R. Bennett, J. Acoust. Soc. Am., **15**, 164 (1944).

Furthermore $p_n(E)$ satisfies a sufficient set of conditions for its expansion in a Fourier series[10] in the interval $-A<E<A$. Therefore:

$$p_n(E) = \sum_{s=-\infty}^{\infty} c_s e^{is\pi E/A}, \qquad -A < E < A \tag{23}$$

$$c_s = \frac{1}{2A}\int_{-A}^{A} p_n(E)e^{-is\pi E/A}dE = \frac{1}{2A}\prod_{m=1}^{n} J_0\left(\frac{a_m s\pi}{A}\right). \tag{24}$$

The last step follows by substituting $t=s\pi/A$ in (22). Hence

$$\begin{aligned} p_n(E) &= \frac{1}{2A}\sum_{s=-\infty}^{\infty} e^{is\pi E/A}\prod_{m=1}^{n} J_0(s\pi a_m/A) \\ &= \frac{1}{2A}\left[1 + 2\prod_{s=1}^{\infty}\cos s\pi E/A \prod_{m=1}^{n} J_0(s\pi a_m/A)\right]. \end{aligned} \tag{25}$$

The probability Ψ_n of finding an instantaneous absolute value exceeding E is given by

$$\begin{aligned} \Psi_n &= 1 - 2\int_0^E p_n(E)dE \\ &= 1 - \frac{E}{A} - \frac{2}{\pi}\sum_{s=1}^{\infty}\frac{1}{s}\sin s\pi E/A \prod_{m=1}^{n} J_0(s\pi a_m/A). \end{aligned} \tag{26}$$

If $a_1=a_2=\cdots=a_n=a$, $\Psi_n(k)$ is the probability of exceeding ka,

$$\begin{aligned} \Psi_n(k) &= 1 - \frac{2}{\pi}\int_0^{\infty} J_0^n(z)\frac{\sin kz}{z}dz \\ &= 1 - \frac{k}{n} - \frac{2}{\pi}\sum_{s=1}^{\infty}\frac{1}{s}J_0^n(s\pi/n)\sin s\pi k/n. \end{aligned} \tag{27}$$

Values of $\Psi_n(k)$ are listed in Table II.

TABLE II

$$\text{Value of } \Psi_n(k) = 1 - \frac{2}{\pi}\int_0^{\infty} J_0^n(z)\frac{\sin k}{z}dz$$

= probability that instantaneous absolute value of n sine waves of equal peaks exceeds k times peak of one.

k	1	2	3	4	5	6	7	8
Ψ_{10}	.6604	.3779	.1832	.07330	.02337	.00563	.00094	.00009

[10] This method of evaluating integrals by Fourier series has also been used by D. G. Kendall, *A summation formula associated with finite trigonometric integrals*, Quart. J. Math. (Ox. Ser.) **13**, 172–184 (1942).

It is convenient to express results in terms of the ratio of the quantity to the root mean square value. Let η represent the root mean square value of the envelope and τ the root mean square instantaneous value of n sine waves with equal amplitude a. Then

$$\eta = a\sqrt{n} \tag{28}$$

$$\tau = a\sqrt{n/2} \tag{29}$$

Fig. 2 shows $\Phi_n(k)$ plotted against $10 \log_{10} k^2/n$, the ratio of ka to η expressed in

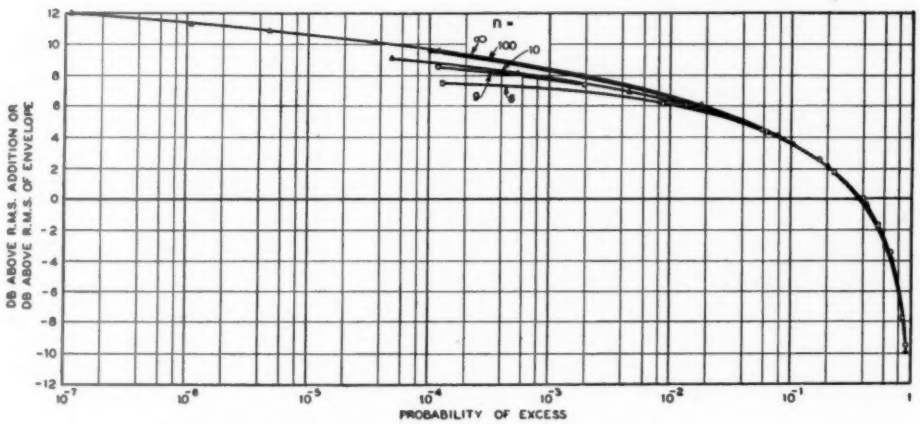


FIG. 2. Distribution function of envelope of n equiamplitude randomly phase sine waves expressed in db above root mean square of envelope.

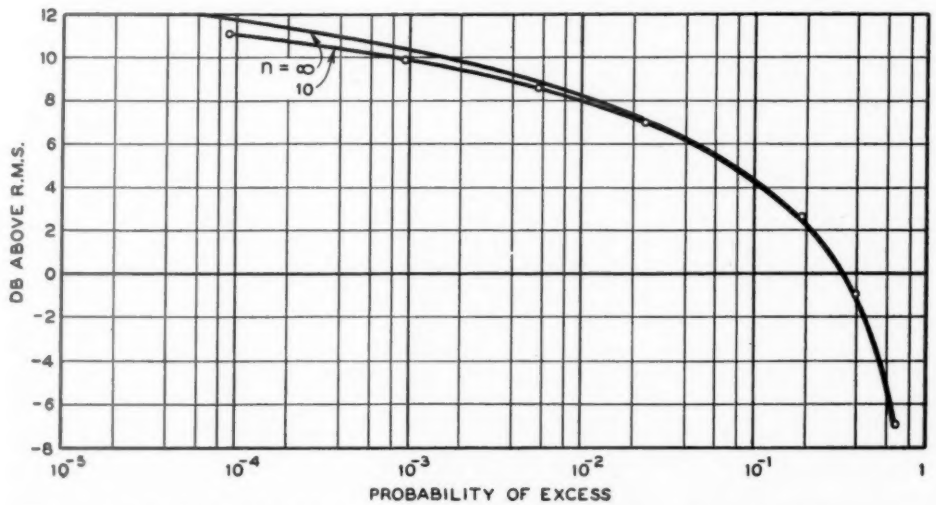


FIG. 3. Distribution function of instantaneous value of n equiamplitude randomly phased sine waves expressed in db above root mean square.

decibels. Likewise Fig. 3 shows $\Psi_n(k)$ plotted against $10 \log_{10} 2k^2/n$, the ratio of ka to τ expressed in decibels. The limiting forms of these curves when n is large[11] are:

$$\lim_{n\to\infty} \Phi_n(k) = e^{-k^2/n} \tag{30}$$

$$\lim_{n\to\infty} \Psi_n(k) = 1 - erf\, k/\sqrt{n}. \tag{31}$$

The latter approaches the normal law in one dimension, while the former approaches the distribution of the radial coordinate in a two-dimensional normal field. The results are in accordance with the central limit theorem.[12] All numerical calculations of this paper were made by Mrs. A. J. Shanklin.

Since completion of this work, a published article[13] has appeared treating the same problems by different methods. Figs. 4 and 5 of the reference are graphs of the function we have called Ψ_n and Figs. 7 and 8 give Φ_n. Agreement between these curves and Tables I and II of the present paper is good except near the maximum amplitudes, where the probabilities plotted on the curves are higher than our tabulated values. The latter have been carefully checked and are believed to be accurate to the number of significant figures listed. Upper bounds have been calculated for the remainder after the actual number of terms used in the series and shown to be negligibly small.

APPENDIX

Envelope of a Group of Sine Waves

Let

$$E = \sum_m E_m \cos p_m t.$$

Then by obvious rearrangements:

$$E = \mathrm{Re} \sum_m E_m e^{ip_m t} = \mathrm{Re}\, e^{ip_0 t} \sum_m e^{i(p_m - p_0)t}$$

$$= \cos p_0 t \sum_m E_m \cos (p_m - p_0)t - \sin p_0 t \sum_m E_m \sin (p_m - p_0)t$$

$$= \left\{ \left[\sum_m E_m \cos (p_m - p_0)t \right]^2 + \left[\sum_m E_m \sin (p_m - p_0)t \right]^2 \right\}^{1/2}$$

$$\cdot \cos \left[p_0 t + \arctan \frac{\sum_m E_m \sin (p_m - p_0)t}{\sum_m E_m \cos (p_m - p_0)t} \right].$$

This is the familiar form of an envelope function multiplying an oscillating function. In order for the resolution to be physically significant, p_0 must be large compared

---

[11] The limiting forms can be derived from the integrals by Laplace's process as described in Watson, *loc. cit.*, Sec. 13.48.

[12] H. Cramér, *loc. cit.* Chap. X, J. V. Uspensky, *Introduction to mathematical probability*, McGraw-Hill, New York, Chap. 15.

[13] Margaret Slack, *The probability distributions of sinusoidal oscillations combined in random phase*, J.I.E.E., **93**, 76–86 (1946).

with $p_m - p_0$. The same expression for the envelope function may be derived from the mathematical definition of an envelope if we regard the defining equation for E as that of a family of curves in E and t with the parameter $(p_m - p_0) = \alpha$. Eliminating α between this equation and the equation obtained by differentiating partially with respect to α gives the square root function appearing above.

The envelope is seen to be the square root of the sum of the squares of the sum of a set of horizontal components $E_m \cos (p_m - p_0)t$ and the sum of a set of vertical components $E_m \sin (p_m - p_0)t$. It is therefore equal to the magnitude of the resultant of a set of two-dimensional vectors with absolute values E_m and phase angles, $(p_m - p_0)t$. If the frequencies $p_1, p_2, \cdots$ are incommensurable, the phases are random and the problem is the same as that of random flights.

We note that the square of the envelope is equal to

$$\sum_m E_m^2 + 2 \sum_{l,m}{}' E_l E_m \cos (p_l - p_m)t$$

where $\sum'$ does not include $l = m$. The average square is therefore $\sum_m E_m^2$.

CONCERNING A NEW TRANSCENDENT, ITS TABULATION AND APPLICATION IN ANTENNA THEORY*

BY

C. J. BOUWKAMP

Natuurkundig Laboratorium der N. V. Philips' Gloeilampenfabrieken, Eindhoven

1. Introduction. As is well known, the integral sine and cosine functions $Si(z)$ and $Ci(z)$, respectively, are frequently met with in problems of applied mathematics. As an example we may mention the theory of antenna radiation, though in this field one preferably uses a slightly different pair of functions $S(z)$, $C(z)$ defined by

$$E(z) = C(z) + iS(z) = \int_0^z (1 - e^{-it})dt/t, \tag{1}$$

wherein i denotes the imaginary unit. Obviously $S(z)$ is identical to $Si(z)$:

$$S(z) = Si(z) = \int_0^z \sin t\, dt/t. \tag{2}$$

Further,[1] if γ denotes Euler's constant,

$$C(z) = \gamma + \log z - Ci(z) = \int_0^z (1 - \cos t)dt/t. \tag{3}$$

Recently, the author was led to the study of another transcendental function closely related to that defined by (1). This new function $E_1(z)$ is related to $E(z)$ in the same way as the latter is to the ordinary trigonometric functions, viz.

$$E_1(z) = \int_0^z E(t)dt/t = \int_0^1 \int_0^1 (1 - e^{-izst})dsdt/st. \tag{4}$$

The function $E_1(z)$ was encountered in antenna theory but may possibly be of some value in other fields as well. Therefore it is thought worth while to treat some of its features here. In addition, a short table of numerical values may be of general interest. Finally, the function $\alpha_2(x)$, as it occurs in Hallén's antenna theory, is shown to be expressible in terms of the functions $E(x)$ and $E_1(x)$.

2. Power series and asymptotic expansion for $E_1(z)$. With respect to numerical evaluation, especially for small values of z, a power-series development may serve the purpose. After expanding the integrand in (1) into powers of t, one simple integration leads to a power series for $E(z)$. Using the latter in the left-hand integral of (4) we obtain, after another term-by-term integration, the required expansion immediately, viz.

$$E_1(z) = -\sum_{n=1}^{\infty} \frac{(-iz)^n}{n^2 \cdot n!}. \tag{5}$$

* Received Nov. 26, 1946.

[1] In our opinion, the short notation $C(z)$ for the integral (3) is to be preferred to those like $\overline{Ci}(z)$ or $Cin(z)$, as suggested by some authors. Then $E(z)$ may be a suitable abbreviation for the combination (1) analogous to the familiar $\exp(iz) = \cos z + i \sin z$.

For large values of z, however, an asymptotic expansion is more desirable. In this respect we have found the following development:

$$E_1(z) \sim A + B \log z + (\log z)^2/2 + \frac{e^{-iz}}{iz} \sum_{n=1}^{\infty} n! \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) (i/z)^n, \quad (6)$$

where the constants A and B are given by

$$A = \gamma^2/2 - \pi^2/24 + \pi\gamma i/2 = -0.24464\ 45548 + 0.90668\ 45943i,$$
$$B = \gamma + \pi i/2 = 0.57721\ 56649 + 1.57079\ 63268i.$$

Formula (6) may be proved as follows: An equivalent definition of $E_1(z)$ is

$$E_1(z) = -\int_0^1 (1 - e^{-izt}) \log t \, dt/t, \quad (7)$$

as can be verified by a partial integration of the left-hand integral in (4), and an obvious change in the variable of integration. Once more integrating by parts one is led to

$$E_1(z) = \frac{1}{2} iz \int_0^1 e^{-izt} (\log t)^2 dt. \quad (8)$$

Now, for large[2] values of z, the main contribution to the integral (8) comes from the values of the integrand in the neighbourhood of $t=0$. It is therefore reasonable to consider the integral (8) as the sum of two terms:

$$E_1(z) = H(z) + h(z), \quad (9)$$

where the "main term" and the "correction term" are defined by

$$H(z) = \frac{1}{2} iz \int_0^{-i\infty} e^{-izt} (\log t)^2 dt, \quad (10)$$

$$h(z) = \frac{1}{2} iz \int_{-i\infty}^1 e^{-izt} (\log t)^2 dt, \quad (11)$$

respectively.

Let us first transform the main term $H(z)$. Evidently one can transform the expression (10) into

$$H(z) = \frac{1}{2} iz \left[\frac{d^2}{ds^2} \int_0^{-i\infty} t^{s-1} e^{-izt} dt\right]_{s=1}.$$

Now, from gamma-function theory, we have

$$\int_0^{-i\infty} t^{s-1} e^{-izt} dt = \Gamma(s)/(iz)^s.$$

Consequently, upon performing the differentiations,

$$H(z) = \Gamma''(1)/2 - (\log z + \pi i/2)\Gamma'(1) + (\log z + \pi i/2)^2 \Gamma(1)/2.$$

[2] Henceforth we suppose $z>0$.

Finally, after substitution of the known numerical constants

$$\Gamma(1) = 1, \qquad \Gamma'(1) = -\gamma, \qquad \Gamma''(1) = \gamma^2 + \pi^2/6,$$

one easily finds

$$H(z) = A + B \log z + (\log z)^2/2,$$

wherein the coefficients A, B are as specified above.

Concerning the correction term $h(z)$ we proceed as follows: Let

$$g(t) = (\log t)^2/2;$$

then, by successive partial integrations of (11),

$$h(z) = e^{-iz}\left[g(1) + \frac{g'(1)}{iz} + \cdots + \frac{g^{(n)}(1)}{(iz)^n}\right] - \frac{1}{(iz)^n}\int_1^{-i\infty} g^{(n+1)}(t)e^{-izt}dt.$$

Further, by induction, or otherwise,

$$g^{(n+1)}(t) = (-1)^n n!\, t^{-n-1}\left(\log t - 1 - \frac{1}{2} - \cdots - \frac{1}{n}\right),$$

$$g(1) = g'(1) = 0, \qquad g^{(n+1)}(1) = (-1)^{n+1}n!\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right).$$

We thus obtain for $h(z)$, after $N = n-1$ terms,

$$h(z) = \frac{e^{-iz}}{iz}\sum_{n=1}^{N} n!\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)(i/z)^n + R_N(z), \tag{12}$$

where the remainder $R_N(z)$ is given by

$$R_N(z) = \frac{i^{N+1}(N+1)!}{z^{N+1}}\int_1^{-i\infty} \frac{e^{-izt}}{t^{N+2}}\left\{1 + \frac{1}{2} + \cdots + \frac{1}{N+1} - \log t\right\}dt. \tag{12a}$$

Moreover, it can be shown that

$$|R_N(z)| \leqq \left(1 + \frac{2/(N+1)}{1 + 1/2 + \cdots + 1/(N+1)}\right) \times |\text{ last term taken into account }|,$$

$$|R_N(z)| \leqq \left(1 + \frac{1/z}{1 + 1/2 + \cdots + 1/(N+1)}\right) \times |\text{ first term not taken into account }|.$$

Further details are left to the reader.

We have thus proved the validity of the asymptotic expansion (6) for positive values of z, though (6) holds for $\mathrm{Re}(z) > 0$ as well.

3. A third development for the function $E_1(z)$. For moderate values of z (for instance $z = 10$) neither the power series (5) nor the asymptotic development (6) is very useful for numerical purposes, as then too many terms are required. For such values of z it is better to apply the Taylor series.

To that end we use the obvious formula

$$E_1(z + \Delta) = E_1(z) + \sum_{n=0}^{\infty} \frac{d^n}{dz^n}\left\{\frac{E(z)}{z}\right\}\frac{\Delta^{n+1}}{(n+1)!} \equiv E_1(z) + \sum_{n=1}^{\infty}\frac{\Delta^n}{n!}c_n(z). \tag{13}$$

The series (13) converges for all values of Δ because $E_1(z+\Delta)$ is an integral function of Δ. The following recurrence relations for the coefficients $c_n(z)$, $(n>0)$, can be established:

$$z^2c_{n+2}(z) + (2n+1)zc_{n+1}(z) + n^2c_n(z) = (-1)^{n+1}i^ne^{-iz}, \tag{14}$$

with initial values

$$c_1(z) = E(z)/z, \qquad c_2(z) = \left[1 - e^{-iz} - E(z)\right]/z^2.$$

Given z, the functions $c_n(z)$ can be calculated successively. As at present there exist very accurate tables for the integral sine and cosine functions,[3] it is not difficult to prepare an auxiliary table for the function $E(z)$.

4. A short table for the function $E_1(z)$. We have prepared a short six-decimal table for $E_1(z)$ for values of z between 0.0 and 20.0 at intervals of length 0.2. For $z \leqq 5.0$ the power series was applied, up to 8 decimals. For $5.0 \leqq z \leqq 20.0$ the function was computed by means of (13), the functions $c_n(z)$ being pre-calculated (8 decimals) for $z = 5, 7, \cdots, 19$. Accordingly, (13) was successively applied with values of $|\Delta|$ not exceeding 1.0.

The value of $E_1(20)$, obtained in this way, was checked by application of the asymptotic series for $z=20$. The difference appeared only two units of the eighth decimal. The values of $E_1(10)$ and $E_1(15)$ were also checked, by comparison with the power-series values. Moreover the eight-decimal numbers on the worksheet were checked by calculating sixth-order differences, and then rounded off to six decimals. Therefore, it will be very unlikely that the error therein exceeds half a unit of the last decimal.

Tables I, II contain the real and imaginary parts of the function $E_1(z)$, respectively. Thus they give

$$\text{Re } E_1(z) = \int_0^z [\gamma + \log t - Ci(t)]dt/t = z\int_0^1 \sin zt\,(\log t)^2\,dt/2, \tag{15}$$

$$\text{Im } E_1(z) = \int_0^z Si(t)dt/t = z\int_0^1 \cos zt\,(\log t)^2\,dt/2. \tag{15a}$$

5. Hallén's second-order function $\alpha_2(x)$. Hallén[4] derived the following expression for the self-impedance of the center-fed perfectly conducting cylindrical antenna

$$Z(x) = -60i\Omega\,\frac{\cos x + \alpha_1(x)/\Omega + \alpha_2(x)/\Omega^2 + \cdots}{\sin x + \beta_1(x)/\Omega + \beta_2(x)/\Omega^2 + \cdots}.$$

In this formula Ω denotes a large constant, depending on the radius a and the half-length l of the antenna: $\Omega = 2\log(2l/a)$. Further $x = kl = 2\pi l/\lambda$, where λ is the wavelength.

Only the first-order coefficients $\alpha_1(x)$ and $\beta_1(x)$ can be given explicitly in terms of known functions, namely

$$\alpha_1(x) = \tfrac{1}{2}e^{ix}E(4x) - \cos xE(2x). \tag{16}$$

$$\beta_1(x) = \tfrac{1}{2}ie^{ix}\left\{E(4x) - 4E(2x)\right\} + \sin x\left\{\log 4 - E(2x)\right\}. \tag{16a}$$

[3] Tables of sine, cosine and exponential integrals, vols. I, II; New York (1940). Table of sine and cosine integrals; New York (1942). (Federal Works Agency, W.P.A., City of New York.)

[4] Erik Hallén, Nova acta reg. soc. sci. Upsaliensis (4) 11, 1044 (1938).

Rather intricate formulae were given for the second-order coefficients $\alpha_2(x)$ and $\beta_2(x)$, which were evaluated by graphical methods.[5] It may be noticed that in the refined theory[6] the same second-order coefficients occur.

Recently $\alpha_2(x)$ was found to be expressible in terms of $E(x)$ and $E_1(x)$ by means of a fairly simple formula, viz.

$$\alpha_2(x) = -\alpha_1(x)\{\log 4 + E(2x)\} - \cos xE^2(2x)/2 + 2i \sin xE_1(4x) + \cos x\{E_1(4x) - 2E_1(2x)\}. \tag{17}$$

With the help of our tables for the function $E_1(x)$, and the American tables for $Si(x)$, $Ci(x)$, we have calculated α_1 and α_2 to six decimals for 0.0(0.1)5.0. After careful checking, these results were rounded off to four decimals. The final data are given in tables III, IV, whereby

$$\alpha_1(x) = \overset{\text{I}}{\alpha_1}(x) + i\overset{\text{II}}{\alpha_1}(x), \qquad \alpha_2(x) = \overset{\text{I}}{\alpha_2}(x) + i\overset{\text{II}}{\alpha_2}(x).$$

Comparison of our table for the first-order coefficient $\alpha_1(x)$ with those of King and Blake[7] shows only small differences in the last decimal. Also the values of the second-order coefficient $\alpha_2(x)$ are in good agreement with the corresponding two-decimal values obtained by graphical integration.[5]

As for the other second-order coefficient, we do not think it possible to express $\beta_2(x)$ in such a simple way; unfortunately, more intricate functions seem to play a part.

In the following sections a proof of formula (17) will be given. As a rather large amount of analysis seems necessary to establish such proof, we may once more emphasize the usefulness of the short abbreviation $E(x)$ as was adopted here for the combined integral sine and cosine functions.

6. Some auxiliary functions. We introduce the following four functions:

$$\phi_1(x) = \int_0^x \frac{E(x) - E(t)}{x - t}\,dt = -\int_0^x \log(1 - t/x)(1 - e^{-it})dt/t, \tag{18}$$

$$\phi_2(x) = \int_0^x E(x - t)(1 - e^{-it})dt/t, \tag{19}$$

$$\phi_3(x) = \int_0^x \{E(x - t) - E(x)\}e^{-it}dt/t, \tag{20}$$

$$\phi_4(x) = \int_0^{2x} \{\cos(x - t) - \cos x\} \log(1 - t/2x)e^{-it}dt/t. \tag{21}$$

Between these functions the following relations exist:

$$2\phi_1(x) + \phi_2(x) = 2E_1(x), \tag{22}$$

$$\phi_1(x) + \phi_2(x) + \phi_3(x) = E^2(x), \tag{23}$$

$$\phi_4(x) = e^{ix}\phi_1(4x)/2 - \cos x\,\phi_1(2x). \tag{24}$$

---

[5] C. J. Bouwkamp, Physica **9**, 609–631 (1942).
[6] R. King and D. Middleton, Quart. Appl. Math. **3**, 302–335 (1945).
[7] R. King and F. G. Blake, Proc. Inst. Radio Engrs. **30**, 335–349 (1942).

Formula (22) is especially noteworthy as it does not seem to be some trivial equality. A straightforward proof of it may be established by expanding both sides into powers of x. Evidently, φ_1 and φ_2 are integral functions; their respective power series converge for all finite values of x. One will get

$$\phi_1(x) = -\sum_{n=1}^{\infty} \frac{(-ix)^n}{n \cdot n!}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right), \tag{18a}$$

$$\phi_2(x) = 2\sum_{n=1}^{\infty} \frac{(-ix)^{n+1}}{(n+1)(n+1)!}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right), \tag{19a}$$

and then (22) follows at once on account of (5).

A detailed proof of (24) only will be given as an example. Firstly,

$$\{\cos(x-t) - \cos x\}e^{-it}/t = -\tfrac{1}{2}e^{ix}(1 - e^{-2it})/t + \cos x\,(1 - e^{-it})/t$$

$$= \frac{d}{dt}\{-\tfrac{1}{2}e^{ix}E(2t) + \cos x\,E(t)\}.$$

Therefore,

$$\phi_4(x) = -\int_0^{2x} \log(1 - t/2x)\,\frac{d}{dt}\left[\tfrac{1}{2}e^{ix}\{E(2t) - E(4x)\} - \cos x\{E(t) - E(2x)\}\right]dt.$$

Secondly, by partial integration,

$$\phi_4(x) = \frac{1}{2}e^{ix}\int_0^{2x} \frac{E(2t) - E(4x)}{t - 2x}\,dt - \cos x\int_0^{x} \frac{E(t) - E(2x)}{t - 2x}\,dt,$$

and, after some trivial transformation and by the use of (18), this reduces to (24).

7. Proof of formula (17). Instead of Hallén's functions $F_n(z)$ we take $f_n(z) = F_n(z/k)$. These functions are recurrently defined by

$$f_0(z) = \cos z,$$

$$f_{n+1}(z) = \{f_n(x) - f_n(z)\}\log(1 - z^2/x^2) + \int_{-x}^{x} \frac{\{f_n(x) - f_n(\zeta)\}\exp(-i|z + \zeta|) - \{f_n(x) - f_n(z)\}}{|z - \zeta|}\,d\zeta.$$

Apart from f_0, only f_1 can be given explicitly in terms of known functions, viz.

$$f_1(z) = (\cos x - \cos z)\log(1 - z^2/x^2) + \tfrac{1}{2}\{e^{iz}E(2x + 2z) + e^{-iz}E(2x - 2z)\} - \cos x\{E(x + z) + E(x - z)\}. \tag{25}$$

The functions $\alpha_1(x)$ and $\alpha_2(x)$ are obtained when $z = x$ is substituted in $f_1(z)$ and $f_2(z)$, respectively. Therefore the required expression for $\alpha_2(x)$ has to be derived from

$$\alpha_2(x) = \int_0^{2x} \{f_1(x) - f_1(x - t)\}\,\frac{e^{-it}}{t}\,dt. \tag{26}$$

We first write

$$f_1(x) - f_1(x - t) = T_1 + T_2 + \cdots + T_7,$$

wherein

$$T_1 = \{\cos(x-t) - \cos x\} \log(2t/x), \tag{27}$$

$$T_2 = -e^{ix}\cdot e^{it}E(2t)/2, \tag{28}$$

$$T_3 = \cos xE(t) \tag{29}$$

$$T_4 = e^{ix}E(4x)(1-e^{-it})/2, \tag{30}$$

$$T_5 = \{\cos(x-t) - \cos x\} \log(1-t/2x), \tag{31}$$

$$T_6 = \cos x\{E(2x-t) - E(2x)\}, \tag{32}$$

$$T_7 = -e^{ix}\cdot e^{-it}\{E(4x-2t) - E(4x)\}/2. \tag{33}$$

Let the corresponding contributions to the integral in (26) be denoted by I_n, thus

$$I_n = \int_0^{2x} T_n e^{-it}dt/t$$

Then one will find consecutively

$$I_1 = -\alpha_1(x)\log 4 + e^{ix}E_1(4x)/2 - \cos xE_1(2x), \tag{27a}$$

$$I_2 = -e^{-ix}E_1(4x)/2, \tag{28a}$$

$$I_3 = \cos x\{E_1(2x) - E^2(2x)/2\} \tag{29a}$$

$$I_4 = e^{ix}E(4x)\{E(4x) - E(2x)\}/2, \tag{30a}$$

$$I_5 = e^{ix}\phi_1(4x)/2 - \cos x\phi_1(2x), \tag{31a}$$

$$I_6 = \cos x\{E^2(2x) - \phi_1(2x) - \phi_2(2x)\}, \tag{32a}$$

$$I_7 = e^{ix}\{\phi_1(4x) + \phi_2(4x) - E^2(4x)\}/2. \tag{33a}$$

It is thought unnecessary to give detailed proofs of the above formulae, as the general lines are the same as in the example of the preceeding section.

Upon substituting (27a) $\cdots$ (33a) in $\alpha_2(x) = I_1 + I_2 + \cdots + I_7$, we obtain

$$\alpha_2(x) = \cos xE^2(2x)/2 - e^{ix}E(2x)E(4x)/2 - \alpha_1(x)\log 4 + i\sin xE_1(4x)$$
$$- \cos x\{2\phi_1(2x) + \phi_2(2x)\} + e^{ix}\{2\phi_1(4x) + \phi_2(4x)\}/2.$$

On account of (22), the functions φ_1 and φ_2 can be eliminated. One then easily obtains the required formula (17).

TABLE I

z	ReE_1	z	ReE_1	z	ReE_1	z	ReE_1
0.0	0.000 000	5.0	1.972 538	10.0	3.726 338	15.0	4.982 566
0.2	0.004 996	5.2	2.066 256	10.2	3.784 619	15.2	5.025 625
0.4	0.019 933	5.4	2.157 214	10.4	3.842 420	15.4	5.068 443
0.6	0.044 664	5.6	2.245 360	10.6	3.899 692	15.6	5.111 035
0.8	0.078 943	5.8	2.330 699	10.8	3.956 381	15.8	5.153 410
1.0	0.122 434	6.0	2.413 282	11.0	4.012 436	16.0	5.195 569
1.2	0.174 714	6.2	2.493 205	11.2	4.067 808	16.2	5.237 508
1.4	0.235 281	6.4	2.570 598	11.4	4.122 452	16.4	5.279 217
1.6	0.303 564	6.6	2.645 618	11.6	4.176 332	16.6	5.320 681
1.8	0.378 933	6.8	2.718 446	11.8	4.229 418	16.8	5.361 884
2.0	0.460 706	7.0	2.789 276	12.0	4.281 693	17.0	5.402 804
2.2	0.548 165	7.2	2.858 310	12.2	4.333 147	17.2	5.443 422
2.4	0.640 563	7.4	2.925 751	12.4	4.383 783	17.4	5.483 715
2.6	0.737 142	7.6	2.991 799	12.6	4.433 614	17.6	5.523 663
2.8	0.837 139	7.8	3.056 642	12.8	4.482 661	17.8	5.563 247
3.0	0.939 800	8.0	3.120 458	13.0	4.530 954	18.0	5.602 453
3.2	1.044 392	8.2	3.183 403	13.2	4.578 533	18.2	5.641 267
3.4	1.150 210	8.4	3.245 618	13.4	4.625 441	18.4	5.679 683
3.6	1.256 591	8.6	3.307 218	13.6	4.671 727	18.6	5.717 694
3.8	1.362 916	8.8	3.368 298	13.8	4.717 442	18.8	5.755 304
4.0	1.468 623	9.0	3.428 929	14.0	4.762 639	19.0	5.792 515
4.2	1.573 207	9.2	3.489 159	14.2	4.807 370	19.2	5.829 339
4.4	1.676 231	9.4	3.549 016	14.4	4.851 684	19.4	5.865 788
4.6	1.777 320	9.6	3.608 507	14.6	4.895 628	19.6	5.901 879
4.8	1.876 168	9.8	3.667 622	14.8	4.939 244	19.8	5.937 632
5.0	1.972 538	10.0	3.726 338	15.0	4.982 566	20.0	5.973 068

TABLE II

z	ImE_1	z	ImE_1	z	ImE_1	z	ImE_1
0.0	0.000 000	5.0	3.467 907	10.0	4.526 334	15.0	5.157 090
0.2	0.199 852	5.2	3.527 976	10.2	4.559 056	15.2	5.178 576
0.4	0.398 818	5.4	3.584 497	10.4	4.590 878	15.4	5.199 864
0.6	0.596 026	5.6	3.637 923	10.6	4.621 791	15.6	5.220 929
0.8	0.790 627	5.8	3.688 683	10.8	4.651 802	15.8	5.241 742
1.0	0.981 811	6.0	3.737 180	11.0	4.680 931	16.0	5.262 280
1.2	1.168 815	6.2	3.783 779	11.2	4.709 208	16.2	5.282 518
1.4	1.350 936	6.4	3.828 813	11.4	4.736 676	16.4	5.302 436
1.6	1.527 537	6.6	3.872 571	11.6	4.763 386	16.6	5.322 019
1.8	1.698 057	6.8	3.915 302	11.8	4.789 398	16.8	5.341 255
2.0	1.862 017	7.0	3.957 213	12.0	4.814 776	17.0	5.360 140
2.2	2.019 023	7.2	3.998 470	12.2	4.839 587	17.2	5.378 671
2.4	2.168 772	7.4	4.039 198	12.4	4.863 898	17.4	5.396 855
2.6	2.311 048	7.6	4.079 485	12.6	4.887 779	17.6	5.414 701
2.8	2.445 729	7.8	4.119 385	12.8	4.911 291	17.8	5.432 223
3.0	2.572 779	8.0	4.158 921	13.0	4.934 494	18.0	5.449 442
3.2	2.692 246	8.2	4.198 089	13.2	4.957 441	18.2	5.466 378
3.4	2.804 259	8.4	4.236 865	13.4	4.980 178	18.4	5.483 057
3.6	2.909 021	8.6	4.275 206	13.6	5.002 741	18.6	5.499 504
3.8	3.006 798	8.8	4.313 058	13.8	5.025 158	18.8	5.515 747
4.0	3.097 916	9.0	4.350 357	14.0	5.047 448	19.0	5.531 813
4.2	3.182 750	9.2	4.387 037	14.2	5.069 623	19.2	5.547 727
4.4	3.261 713	9.4	4.423 033	14.4	5.091 683	19.4	5.563 513
4.6	3.335 250	9.6	4.458 283	14.6	5.113 624	19.6	5.579 192
4.8	3.403 823	9.8	4.492 731	14.8	5.135 432	19.8	5.594 782
5.0	3.467 907	10.0	4.526 334	15.0	5.157 090	20.0	5.610 298

TABLE III

x	α_1^{I}	α_1^{II}	x	α_1^{I}	α_1^{II}
0.0	0.0000	0.0000	2.5	0.2361	1.4528
0.1	−0.0100	0.0007	2.6	0.3550	1.3695
0.2	−0.0393	0.0053	2.7	0.4686	1.2690
0.3	−0.0865	0.0175	2.8	0.5756	1.1534
0.4	−0.1490	0.0407	2.9	0.6749	1.0247
0.5	−0.2235	0.0773	3.0	0.7661	0.8851
0.6	−0.3061	0.1292	3.1	0.8487	0.7362
0.7	−0.3924	0.1973	3.2	0.9225	0.5800
0.8	−0.4780	0.2815	3.3	0.9875	0.4180
0.9	−0.5583	0.3807	3.4	1.0436	0.2514
1.0	−0.6291	0.4930	3.5	1.0905	0.0816
1.1	−0.6866	0.6156	3.6	1.1280	−0.0904
1.2	−0.7278	0.7451	3.7	1.1556	−0.2634
1.3	−0.7503	0.8777	3.8	1.1730	−0.4363
1.4	−0.7528	1.0091	3.9	1.1795	−0.6080
1.5	−0.7345	1.1351	4.0	1.1743	−0.7771
1.6	−0.6957	1.2517	4.1	1.1569	−0.9422
1.7	−0.6376	1.3550	4.2	1.1265	−1.1016
1.8	−0.5618	1.4419	4.3	1.0828	−1.2535
1.9	−0.4708	1.5097	4.4	1.0253	−1.3960
2.0	−0.3672	1.5563	4.5	0.9542	−1.5271
2.1	−0.2540	1.5805	4.6	0.8695	−1.6446
2.2	−0.1343	1.5819	4.7	0.7720	−1.7468
2.3	−0.0109	1.5604	4.8	0.6625	−1.8315
2.4	0.1134	1.5170	4.9	0.5423	−1.8973
2.5	0.2361	1.4528	5.0	0.4130	−1.9426

TABLE IV

x	α_2^{I}	α_2^{II}	x	α_2^{I}	α_2^{II}
0.0	0.0000	0.0000	2.5	−2.5787	3.5091
0.1	−0.0359	0.0022	2.6	−2.2678	3.4640
0.2	−0.1415	0.0171	2.7	−1.9045	3.3755
0.3	−0.3099	0.0563	2.8	−1.4931	3.2385
0.4	−0.5311	0.1284	2.9	−1.0397	3.0487
0.5	−0.7919	0.2392	3.0	−0.5524	2.8034
0.6	−1.0784	0.3901	3.1	−0.0402	2.5018
0.7	−1.3762	0.5790	3.2	0.4869	2.1447
0.8	−1.6723	0.8004	3.3	1.0187	1.7349
0.9	−1.9560	1.0462	3.4	1.5450	1.2770
1.0	−2.2191	1.3070	3.5	2.0560	0.7771
1.1	−2.4564	1.5727	3.6	2.5430	0.2427
1.2	−2.6654	1.8344	3.7	2.9979	−0.3177
1.3	−2.8455	2.0844	3.8	3.4141	−0.8948
1.4	−2.9977	2.3171	3.9	3.7862	−1.4792
1.5	−3.1232	2.5291	4.0	4.1100	−2.0611
1.6	−3.2229	2.7192	4.1	4.3828	−2.6312
1.7	−3.2969	2.8875	4.2	4.6030	−3.1808
1.8	−3.3437	3.0355	4.3	4.7700	−3.7021
1.9	−3.3604	3.1644	4.4	4.8840	−4.1880
2.0	−3.3431	3.2752	4.5	4.9459	−4.6330
2.1	−3.2867	3.3680	4.6	4.9568	−5.0323
2.2	−3.1862	3.4413	4.7	4.9180	−5.3825
2.3	−3.0369	3.4923	4.8	4.8308	−5.6811
2.4	−2.8350	3.5166	4.9	4.6959	−5.9264
2.5	−2.5787	3.5091	5.0	4.5142	−6.1172

THE CYLINDRICAL ANTENNA WITH GAP

BY

RONOLD KING AND T. W. WINTERNITZ

Cruft Laboratory and The Lyman Laboratory of Physics, Harvard University

1. Introduction. In the King-Middleton theory of the center-driven cylindrical antenna[1] current and impedance are defined in terms of a discontinuity in scalar potential. Such a driving mechanism is unavailable physically, since actual potential differences necessarily exist across finite distances. In practice, an antenna usually is connected to the conductors of a transmission line across which a potential difference is maintained by a generator elsewhere along the line. A common arrangement is shown in Fig. 1. The antenna in Fig. 1 differs from a conductor center-driven by a discontinuity in scalar potential in several respects. First, there is a gap in place of a

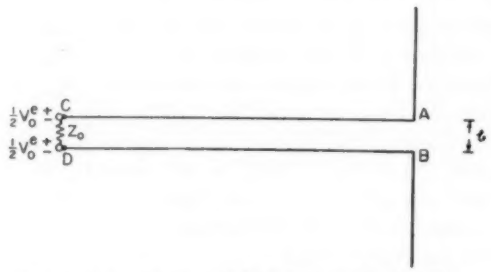


FIG. 1. Antenna center-driven from a two-wire line.

current-carrying section of conductor in the region between A and B or $z=-\delta$ and $z=\delta$. Second, a part of the space around the antenna is occupied by the two conductors of the transmission line carrying equal and opposite distributions of current and charge separated a *finite* distance $b \doteq 2\delta$. Third, there exist small junction regions (of magnitude comparable with the radius a of the conductors) where each half of the antenna joins one of the conductors of the line. These regions are not obviously a part of either the antenna alone or the line alone, and thus introduce an element of vagueness as to where the antenna ends and the line begins.

The rigorous analysis of the complete circuit of Fig. 1 (as of most other electric networks including even a simple series circuit consisting of a coil and a condenser) is excessively difficult if formulated as a three-dimensional boundary-value problem in electromagnetic field-theory. Fortunately, it is possible to gain much in simplicity at the price of a measure of rigor by requiring the radius a of all conductors to be sufficiently small to satisfy the inequality,

$$\beta_0 = \omega a/v_0 = 2\pi a/\lambda_0 \ll 1. \tag{1}$$

Subject to (1), it is a good approximation to assume that cross-sectional and axial distributions of current in conductors are independent, and so reduce the analysis of complex configurations to one-dimensional form.

[1] R. King and D. Middleton, Quart. Appl. Math. **3**, 302 (1946).

2. One dimensional theory. The general formulation of one-dimensional electric circuit and antenna theory is available in the literature.[1,2,3,4,5] It is based on the fundamental relations,

$$-\operatorname{grad}\varphi - j\omega\mathbf{A} = \mathbf{E}, \tag{2}$$

$$\operatorname{div}\mathbf{A} + j\frac{\beta_0^2}{\omega}\varphi = 0, \tag{3}$$

where $\mathbf{A}$ is the Lorentz vector potential defined in general by (3) in conjunction with

$$\operatorname{curl}\mathbf{A} = B. \tag{4}$$

A particular integral of the non-homogeneous wave equation in $\mathbf{A}$ obtained by substituting (2), (3), (4) into the Maxwell equation is

$$\mathbf{A} = \frac{1}{4\pi\nu_0}\left\{\int_\tau \frac{\mathbf{i}'}{R}e^{-j\beta_0 R}d\tau' + \int_\Sigma \frac{\mathbf{I}'}{R}e^{-j\beta_0 R}d\sigma'\right\}. \tag{5}$$

where $\nu_0 = 1/\mu_0 = 10^7/4\pi$ meters/henry, $\mathbf{i}$ is the volume density of current, and $\mathbf{I}$ the surface density of current. R is the distance from the unprimed coordinate of the point where $\mathbf{A}$ is calculated to the primed coordinates locating the elements of integration. In good conductors, only $\mathbf{i}$ is used; in *perfect* conductors, $\mathbf{i}$ is zero and $\mathbf{I}$ is required. Integration is carried out over *all* conductors of the circuit so that the vector potential $\mathbf{A}$ is uniquely defined by (5) and the scalar potential φ by (3).

For simplicity, let it be assumed that the dielectric is air and that the conductors may be treated as perfect in writing the following boundary conditions for the tangential component of the electric field:

$$\frac{\partial\varphi}{\partial z} + j\omega A_z = -E_z \doteq 0 \text{ on the surface of the antenna;} \tag{6}$$

$$\frac{\partial\varphi}{\partial x} + j\omega A_x = -E_x \doteq 0 \text{ on the surface of each line conductor.} \tag{7}$$

Since there are no currents in the y-direction, A_y vanishes everywhere. Hence, (3) gives

$$\frac{\partial A_z}{\partial z} + \frac{\partial A_x}{\partial x} + j\frac{\beta_0^2}{\omega}\varphi = 0 \tag{8}$$

on the antenna *and* on the line. Differentiation of (8) with respect to z and substitution from (6) gives the following equation:

$$\frac{\partial^2 A_z}{\partial z^2} + \frac{\partial^2 A_x}{\partial z\partial x} + \beta_0^2 A_z = 0 \tag{9}$$

---

[1] R. W. P. King, *Electromagnetic engineering*, McGraw-Hill Book Co., New York, 1945, Vol. 1, Ch. VI and pp. 193–195.

[2] J. R. Carson, Bell Syst. Tech. Journal **6**, No. 1 (1927).

[3] A. T. Starr, *Electric circuits and wave filters*, Pitman Publishing Corp., New York, 1938, Ch. II, especially pp. 41–53.

[4] J. Aharoni, *Antennae*, Oxford Univ. Press, 1946, Ch. II.

on the antenna. Similarly, differentiation of (8) with respect to x and substitution from (7) gives

$$\frac{\partial^2 A_x}{\partial x^2} + \frac{\partial^2 A_z}{\partial x \partial z} + \beta_0^2 A_x = 0 \tag{10}$$

on each conductor of the line.

Since it has been assumed in (1) that all conductors have a radius a that is sufficiently small so that $\beta_0 a \ll 1$ is a good approximation, the contributions to the vector potential at all outside points (including the surface of the conductor) due to the radial components of current in that conductor may be neglected compared with the contributions due to the axial components of current. This is equivalent to assuming that the currents in the antenna set up only a z-component of the vector potential, the currents in the transmission line only an x-component of **A**. It has been shown[6,7] that if the current in a cylindrical conductor has rotational symmetry and the conductor is long compared with its radius, the vector potential on the surface of the conductor differs negligibly from the value it would have if the total current in the conductor were concentrated on the axis instead of being distributed in the manner predicted by skin-effect theory. That is, the component of the Lorentz potential on the surface of the antenna due to the axial currents in the antenna is given by

$$A_z = \frac{1}{4\pi\nu_0}\left\{\int_{-h}^{-s} \frac{I_z'}{R_a} e^{-j\beta_0 R_a} dz' + \int_{s}^{h} \frac{I_z'}{R_a} e^{-j\beta_0 R_a} dz'\right\}. \tag{11}$$

Similarly, on the surface of one of the conductors of the line, the component of the Lorentz potential due to the equal and opposite currents in both conductors is given by:

$$A_x = \frac{1}{4\pi\nu_0}\int_0^s I_x' \left[\frac{e^{-j\beta_0 R_a}}{R_a} - \frac{e^{-j\beta_0 R_b}}{R_b}\right] dx'. \tag{12}$$

Since A_x is determined entirely from currents in the line and A_z entirely from currents in the antenna, the terms with mixed derivatives in (9) and (10) represent coupling between line and antenna. This is significant within distances from A and B (Fig. 1) that are *not* great compared with b. At distances from A and B which are great compared with b, these terms are both negligible and the two equations reduce to

$$\frac{\partial^2 A_z}{\partial z^2} + \beta_0^2 A_z = 0 \text{ on the surface of the antenna,} \tag{13}$$

$$\frac{\partial^2 A_x}{\partial x^2} + \beta_0^2 A_x = 0 \text{ on the surface of each conductor of the line.} \tag{14}$$

---

[6] Ref. 2, pp. 239–243.

[7] O. Zinke, Archiv für Elektrotechnik, **35**, 67 (1941). This work of Zinke shows that the contributions to A_z at z due to currents in different elements dz' at z' depend upon an effective radius r_w that is smaller than a for $|z-z'| < 0.4a$ and larger than a for $|z-z'| > 0.4a$. The value of $\int_{-s}^{s} dz'/R_w$ where

$$R_w = \sqrt{(z-z')^2 + r_w^2}$$

differs from $\int_s^s dz'/R_a$, where $R_a = \sqrt{(z-z')^2 + a^2}$ by less than 1% if $s \geqq 2.5a$ for high frequency values of skin depth in copper.

Of these equations, the first is the one solved in the King-Middleton analysis of the cylindrical antenna[1] by an extension of the method of Hallén;[8] the second is that used in the fundamental derivation of the conventional transmission-line equations for current and voltage.[9]

The derivation of the conventional transmission-line equation for current from (14) involves certain approximations since the current does *not* actually satisfy the transmission-line equation *exactly*, even for perfect conductors, as do both the vector and scalar potentials. The current can be made to satisfy the transmission-line equations approximately only if the effect of radiation on the distribution of current is made negligible by imposing the inequality,

$$\beta_0 b \ll 1, \tag{15}$$

and if, in addition, it is a good approximation to set

$$I_x = [(A_x)_1 + (A_x)_2]/l_0^e(x) \tag{16}$$

where

$$l_0^e(x) = \frac{1}{\pi\nu_0}\left[\ln\frac{b}{a} - \frac{1}{2}\ln F(x)\right] \doteq \frac{1}{\pi\nu_0}\ln\frac{b}{a} = l_0^e. \tag{17}$$

The function $F(x)$ reduces practically to unity[10] at distances from the ends of the line which are great compared with b. Only when this is true, i.e., when

$$(s - x)^2 \gg b^2; \qquad x^2 \gg b^2; \tag{18}$$

does (17) give the *conventional* external inductance per unit length of line, l_0^e. On the other hand, at the ends where $x=0$ or s, $F(x)=b/a$, so that $l_0^e(x)$ is only one half of l_0^e, its normal value.[11] Correspondingly, the capacitance per unit length, $c=\epsilon_0\mu_0/l_0^e(x)$, becomes double the value far from the ends. Only when both (15) and (18) are satisfied does the current obey the transmission-line equations to a good approximation.

Summarizing, it is clear that subject to $\beta_0 a < \beta_0 b \ll 1$ the conventional transmission-line equations in the total current and the potential difference are good approximations only at distances from the ends of the line that are reasonably large compared with the distance b between conductors. At such distances the line parameters are quite constant and, in addition, the coupling between line and antenna due to the mixed derivative terms in (9) and (10) is negligible.

From the point of view of a three-dimensional boundary value problem in field theory in which the boundary conditions of the terminated line are satisfied using an expansion in terms of characteristic wave functions, the conditions (18) that make the conventional line equations good approximations are equivalent to specifying distances from the ends of the line that are sufficiently great to make all exponentially damped higher modes negligible. Current and voltage are then represented to a good approximation by the dominant transmission-line mode.[12]

---

[8] E. Hallén, *Theoretical investigations into the transmitting and receiving qualities of antennae*, Nova Acta Uppsala, (4) **11**, No. 4 (1938).

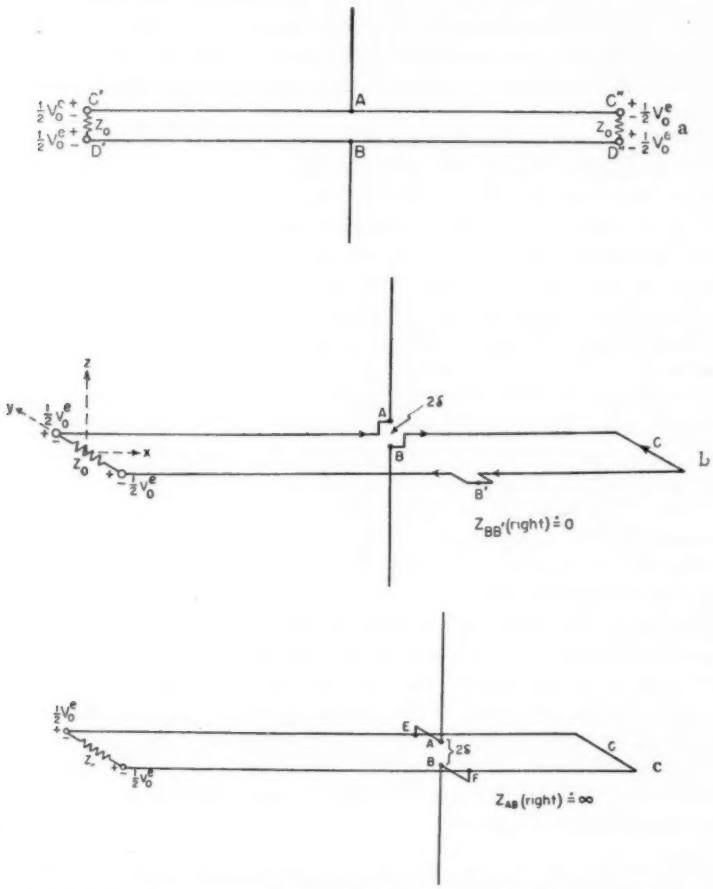
[9] See for example, Ref. 2, pp. 468–477.

[10] Ref. 2, p. 473, Eq. 185.

[11] E. M. Siegel, Univ. of Texas Publ., No. 4031 (1940).

[12] J. R. Carson, A.I.E.E. Journal, **43**, 908 (1924).

3. Definition of impedance. Since the antenna in the circuit of Fig. 1 has been shown to be coupled to the adjacent parts of the transmission line, it is not possible to define for the antenna an impedance $Z_{AB}=Z_\delta$ that is a function only of its own physical properties. It is, of course, possible to define the impedance of the antenna and an attached short section of transmission line that is sufficiently long so that coupling to the termination and end effects are negligible. The impedance so defined



FIGS. 2a, b, c. Antenna center-driven in different ways from a two-wire line.

may then be transformed to the actual location of the antenna using conventional line theory. This is the procedure commonly used both for wave guides and conventional lines in determining equivalent impedances or equivalent lumped-constant networks for discontinuities or terminations. Evidently an impedance so determined is not a characteristic property of the termination alone, e.g. the antenna in Fig. 1, but is that *fictitious* impedance that would have to terminate a *particular* line *if* this had the physically impossible property of having the conventional line equations everywhere

rigorously true, including regions near the junction with the termination. The investigation of such equivalent or apparent impedances for an antenna driven in various ways from two-wire and coaxial lines is in progress both theoretically and experimentally. Three arrangements for a two-wire line different from that in Fig. 1 are shown in Figs. 2a, b, c. A moment's reflection suggests that the mixed derivative terms in (9) and (10), as well as the ratio $[(A_z)_1+(A_z)_2]/I_z$ may have quite different values in the four circuits in Figs. 1 and 2 near the antenna. Therefore, it is not to be expected that the equivalent dominant mode impedances will be the same in the four cases even when the antennas themselves and the gap between their halves are made identical. The quantitative effect of these special differences is not the subject of this paper.

The present purpose is to generalize the King-Middleton theory as much as possible without limiting it specifically to one of the circuits shown in Figs. 1 and 2, or to any other. What can be determined in general is the effect on the current and impedance of a cylindrical antenna with hemispherical caps of a gap of length 2δ between the halves of the antenna, across which a scalar potential difference V_δ is maintained in an unspecified manner. The driving potential difference is to be measured between adjacent rings on the cylindrical surfaces of the two halves of the antenna. Depending upon the nature of the actual driving circuit, there may be exposed flat-end surfaces if the cylinders are solid, or inner surfaces and sharp edges if they are tubes, or elbows, or other junction regions. These are not included in the analysis. In simple cases, they may constitute a capacitance in parallel with the antenna, or they may be part of a four-terminal network, or account may be taken of them in other ways in the analysis of an actual circuit such as is shown in Figs. 1 and 2. In any event, they constitute a part of the problem of a *particular complete* circuit, not of the *general* effect of a separation of the halves of a symmetrical antenna insofar as the interaction of currents and charges in its *cylindrical* surfaces are concerned. These alone are considered here.

The impedance calculated in the next section is, by definition, the ratio of the scalar potential difference maintained across equipotential rings separated a distance 2δ and the total axial current entering the antenna across planes bounded by these rings. In good conductors, most of the current is in a relatively thin layer near the surface so that, as shown by Zinke,[7] the entire current may be assumed concentrated at one half the skin depth. (Note that the vector potential on the surface of the conductor is always defined *outside* the region of non-vanishing current density.) The impedance is defined as follows:[13]

[13] A definition of impedance in terms of scalar potential difference instead of the line integral of the electric field as used by other writers[14,15,16,17] is preferred because the scalar potential difference is uniquely defined, whereas the line integral of the electric field depends upon the path of integration. The line integral of the electric field coincides with the scalar potential difference only when a path is chosen along which ω times the line integral of the vector potential vanishes or is negligible, as along a radial path in transmission lines *sufficiently far from the termination.* It is difficult to see how in conventional one-dimensional circuits involving conductors of small cross-section the impedance of a part of a circuit connected to terminals that are separated a finite distance can be defined consistently except in terms of the scalar potential difference maintained across those terminals. For example, the reactive drop across a loosely wound coil of a few turns of wire is not correctly given by the line integral of the electric field *along any path.* This is most easily seen if the coil degenerates into a straight conductor when the line integral of the electric field along the conductor does not include the inductive reactive drop at all, although it is the

$$Z_\delta = \frac{V_\delta}{I_\delta}, \qquad V_\delta = \varphi(\delta) - \varphi(-\delta). \tag{19}$$

4. Theoretical analysis. The integral equation for the current in the cylindrical antenna which has its halves separated by a gap 2δ differs from the former Eq. (8), Ref. 1, primarily in the limits of integration and in the phase of the cosine function. Thus, the equation for the upper half of the antenna is:

$$4\pi\nu_0 A_z = \int_{-h}^{-\delta} I_{z'} R^{-1} e^{-j\beta_0 R} dz' + \int_{\delta}^{h} I_{z'} R^{-1} e^{-j\beta_0 R} dz'$$

$$= \frac{j4\pi}{R_c \cos\beta_0\delta}\left[C_1 \cos\beta_0(z-\delta) + \tfrac{1}{2}V_\delta \sin\beta_0 z - z^i \cos\beta_0\delta \int_\delta^z I(s)\sin\beta_0(z-s)ds\right]$$

$$(\delta \leqq z \leqq h). \tag{20}$$

The following symmetry conditions are satisfied:

$$I(-z) = I(z); \qquad A_z(-z) = A_z(z). \tag{21}$$

For perfect conductors $z^i = 0$. The solution of (20) is carried out by the same method of successive approximations. The mth order current corresponding to (24) of Ref. 1, and to which it reduces with $\delta = 0$, is

$$(I_z)_m = \frac{j2\pi V_\delta}{R_c\psi} \cdot \frac{\sin\beta_0(h-z) + \sum_{n=1}^{m} M_n^\delta(z)/\psi^n}{\cos\beta_0(h-\delta) + \cos\beta_0\delta\sum_{n=1}^{m} F_n^\delta(h)/\psi^n + \sin\beta_0\delta\sum_{n=1}^{m} G_n^\delta(h)/\psi^n} \tag{22}$$

$$(\delta \leqq z \leqq h).$$

The functions $M_n^\delta(z)$, $F_n^\delta(h)$, and $G_n^\delta(h)$ are like the functions $M_n(z)$, $F_n(h)$, and $G_n(h)$ previously defined but with limits of integration from $-h$ to $-\delta$, δ to h, instead of from $-h$ to h.

The mth order impedance corresponding to (29) in Ref. 1 is

$$(Z_\delta)_m = \frac{jR_c\psi}{2\pi} \cdot \frac{\cos\beta_0(h-\delta) + \cos\beta_0\delta\sum_{n=1}^{m} F_n^\delta(h)/\psi^n + \sin\beta_0\delta\sum_{n=1}^{m} G_n^\delta(h)/\psi^n}{\sin\beta_0(h-\delta) + \sum_{n=1}^{m} M_n^\delta(\delta)/\psi^n}. \tag{23}$$

---

principle contribution to the impedance if the conductor is copper. On the other hand, if the scalar potential difference is used both ohmic resistance and inductive reactance drops are included.

In this connection, it is well to note that the field impedances in wave-guide theory defined in terms of line integrals of the electric field are not simply generalizations of circuit impedances in one-dimensional theory. Rather, they are quantities that have properties *analogous* to those of circuit impedances but in terms of quite different and much more arbitrary variables than are scalar potential difference and total current in conductors of small cross-section.

[14] J. A. Stratton and L. J. Chu, J. Appl. Phys. **12,** 230 (1941).

[15] S. A. Schelkunoff, Proc. I.R.E., **29,** 493 (1941).

[16] L. Infeld, Univ. of Toronto, *Radiation Theory,* Report No. 2 (1942).

[17] J. L. Synge, Univ. of Toronto, *Radiation Theory,* Report No. 4 (1942).

In so far as the formal theoretical analysis is concerned, there is no restriction regarding the gap length 2δ. In practice, one-dimensional circuit impedances are useful primarily for terminals that are close together compared with the wavelength. Hence, a reasonable restriction on the gap is

$$\beta_0\delta \ll 1. \tag{24}$$

Subject to (24) it is possible to expand the current (22) in a Maclaurin series in powers of the small quantity $\beta_0\delta$. For sufficiently small values only the first two terms are required, viz.,

$$I_z(\beta_0\delta) \doteq [I_z(\beta_0\delta)]_{\delta=0} + \beta_0\delta\left[\frac{\partial I_z(\beta_0\delta)}{\partial\beta_0\delta}\right]_{\delta=0}. \tag{25}$$

The current at the driving-point is at $z=\delta$. Hence, setting $z=\delta$ in (25),

$$I_\delta(\beta_0\delta) = I_0(0) + \beta_0\delta\left[\frac{\partial I_\delta(\beta_0\delta)}{\partial\beta_0}\right]_{\delta=0}. \tag{26}$$

In (26), $I_0(0)$ is the driving point current calculated by King and Middleton and given in Ref. 1 by equation (24). Thus,

$$I_\delta(\beta_0\delta) = I_0(0)[1-\epsilon], \tag{27}$$

where

$$\epsilon \equiv \epsilon'' - j\epsilon' \equiv -\frac{\beta_0\delta}{I_0(0)}\left[\frac{\partial I_\delta(\beta_0\delta)}{\partial\beta_0\delta}\right]_{\delta=0}. \tag{28}$$

In expanded form,[18]

$$\begin{aligned}\epsilon = \beta_0\delta\Bigg\{&\frac{\sin\beta_0h + G_1(h)/\psi + G_2(h)/\psi^2 + \cdots}{\cos\beta_0h + F_1(h)/\psi + F_2(h)/\psi^2 + \cdots}\\ &+\frac{2}{\psi}\cdot\frac{e^{-j\beta_0h}}{\beta_0h}\cdot\frac{(1-\cos\beta_0h) + [F_1(0)-F_1(h)]/\psi + \cdots}{\cos\beta_0h + F_1(h)/\psi + F_2(h)/\psi^2 + \cdots}\\ &+\frac{2}{\psi}\cdot\frac{e^{-j\beta_0h}}{\beta_0h}\cdot\frac{\sin\beta_0h + G_1(h)/\psi + \cdots}{\sin\beta_0h + M_1(0)/\psi + M_2(0)/\psi^2 + \cdots}\Bigg\}.\end{aligned} \tag{29}$$

The impedance (23) may be obtained from (25) by noting that

$$Z_\delta/Z_0 = I_0(0)/I_\delta(\delta) = 1/(1-\epsilon), \tag{30}$$

so that,

$$R_\delta = \frac{R_0(1-\epsilon'') + X_0\epsilon'}{(1-\epsilon'')^2 + (\epsilon')^2}, \tag{31}$$

$$X_\delta = \frac{X_0(1-\epsilon'') - R_0\epsilon'}{(1-\epsilon'')^2 + (\epsilon')^2}. \tag{32}$$

[18] The differentiation is straightforward; use is made of the relation $F_1(z)\sin\beta_0h - G_1(z)\cos\beta_0h = 0$ with $F_1(z)$ and $G_1(z)$ as in Ref. 1.

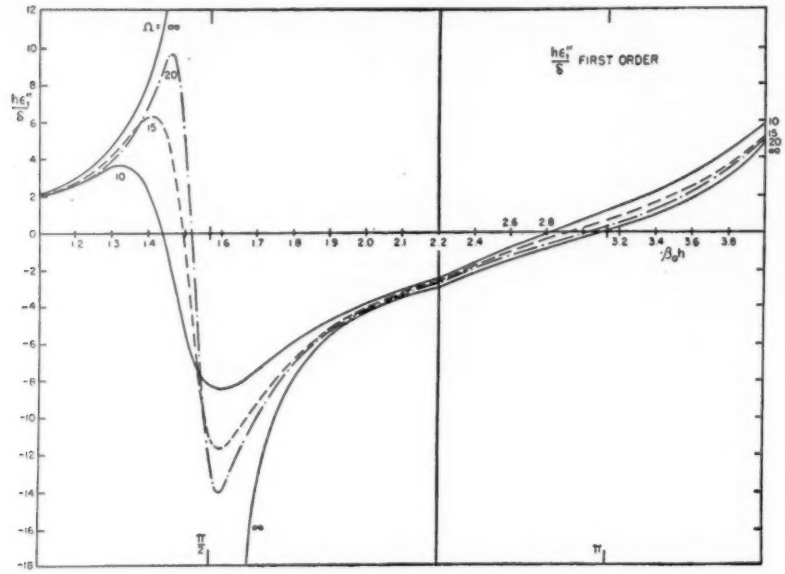


FIG. 3. First order correction factor ϵ_1''.

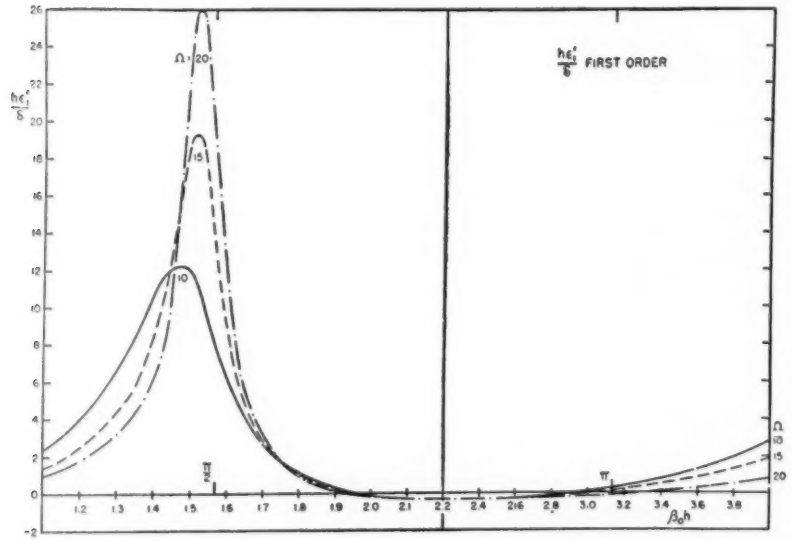


FIG. 4. First order correction factor ϵ_1'.

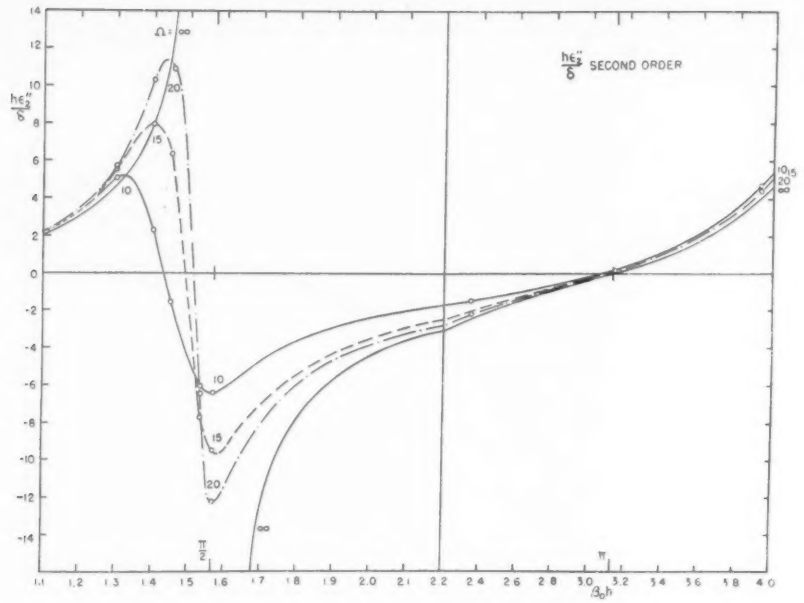


FIG. 5. Second order correction factor ϵ_2''. Calculated points are shown; extrapolation below $\beta_0 h = 1.3$ is estimated using first order curves.

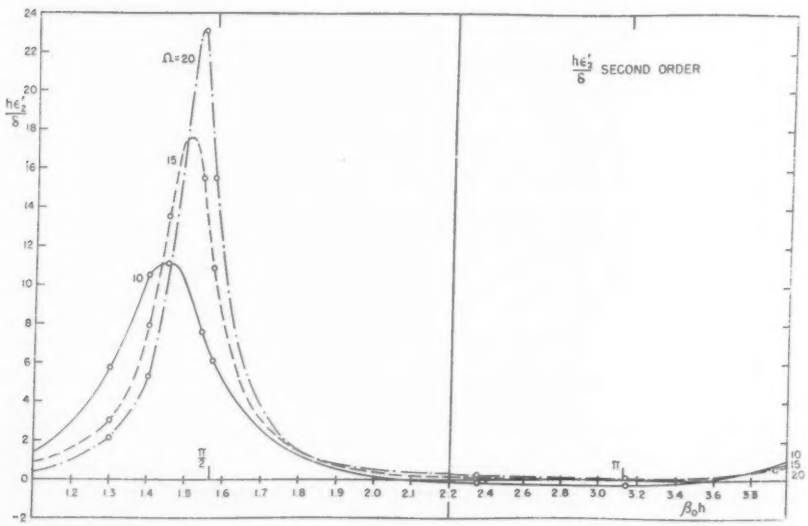


FIG. 6. Second order correction factor ϵ_2'. Calculated points are shown; extrapolation below $\beta_0 h$ is estimated using first order curves.

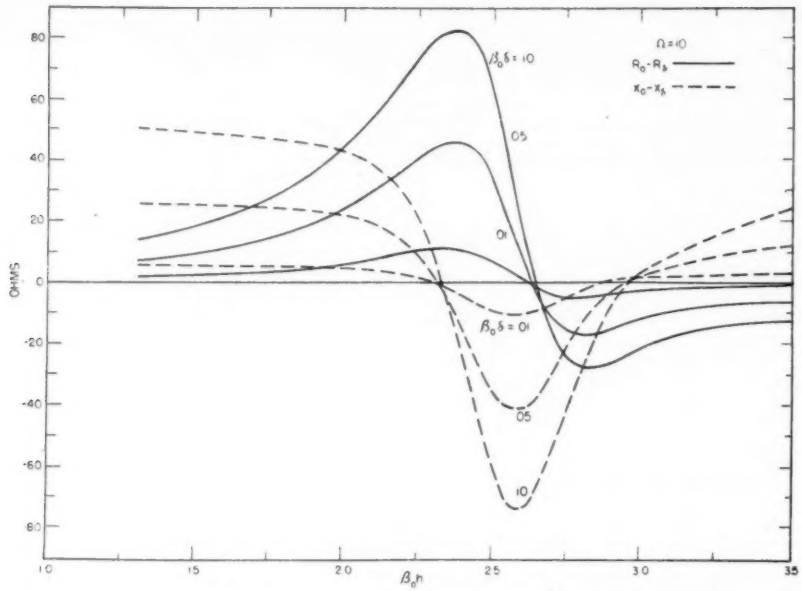


FIG. 7. Second order R_0-R_δ, X_0-X_δ for several values of $\beta_0\delta$ with $\Omega=2ln(2h/a)=10$.

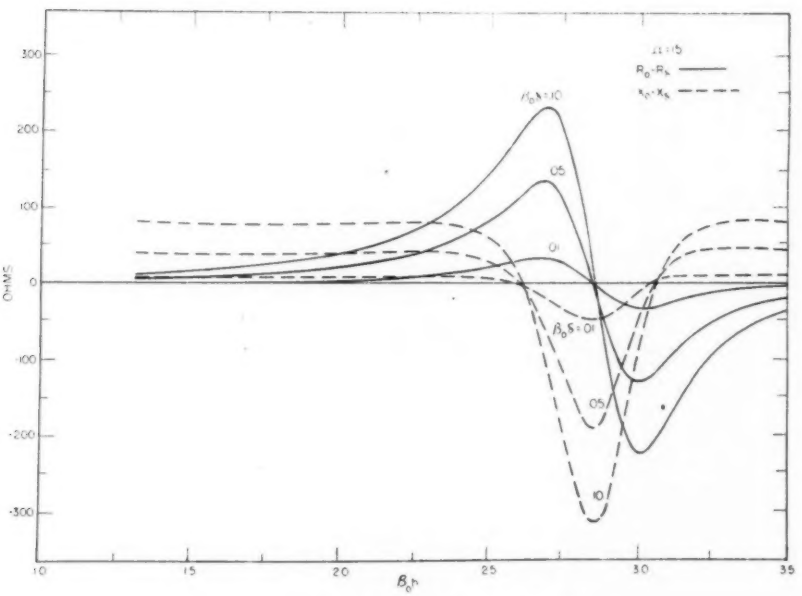


FIG. 8. Same as Fig. 7 with $\Omega=15$.

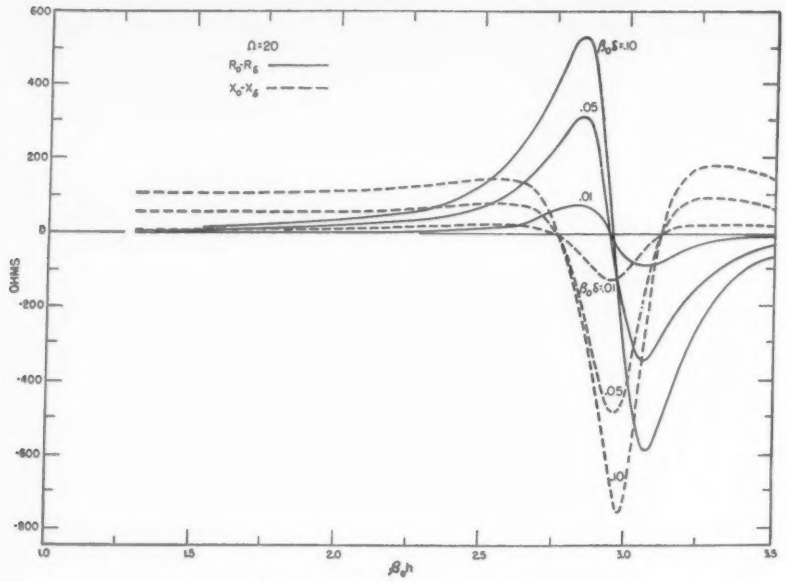


FIG. 9. Same as Fig. 7 with $\Omega=20$.

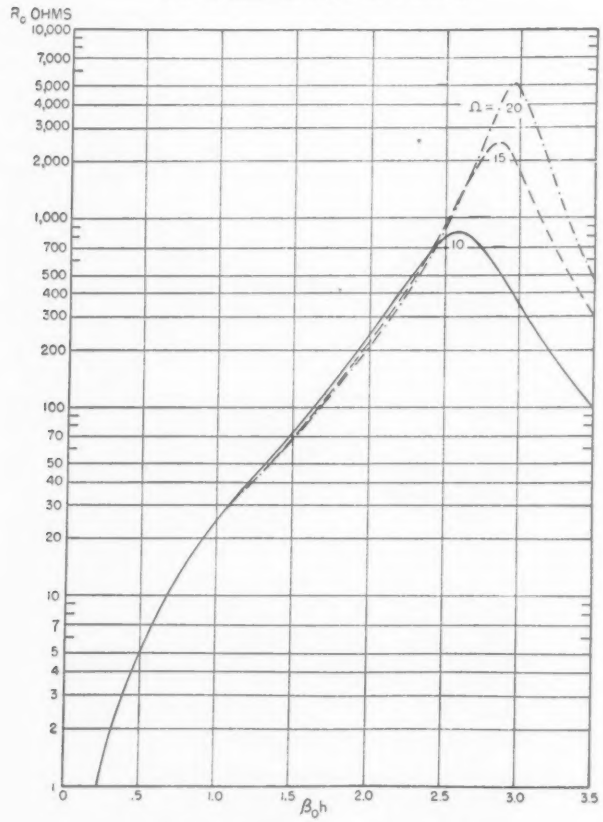


FIG. 10. Second order R_0 as a function of $\beta_0 h$.

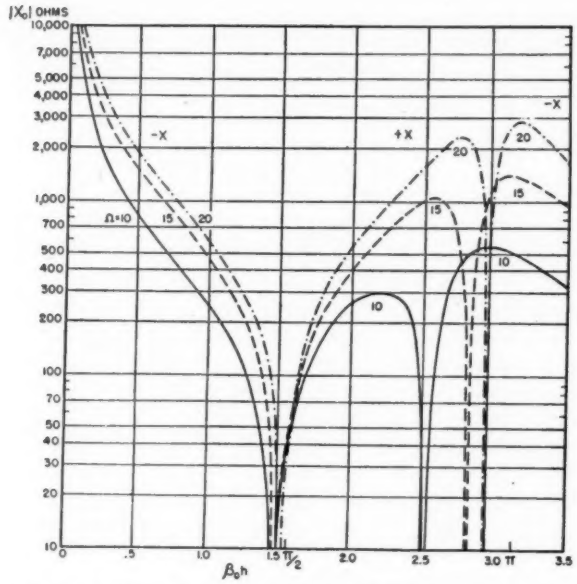


FIG. 11. Second order X_0 as a function of $\beta_0 h$. (Small errors in the curves of Figs. 17–19 in Ref. 1 have been corrected in these curves.)

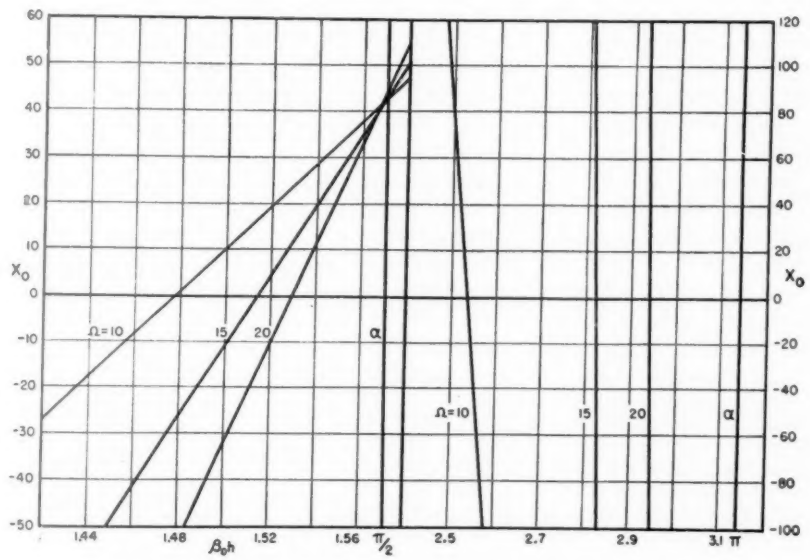


FIG. 12. Second order X_0 near resonance and anti-resonance.

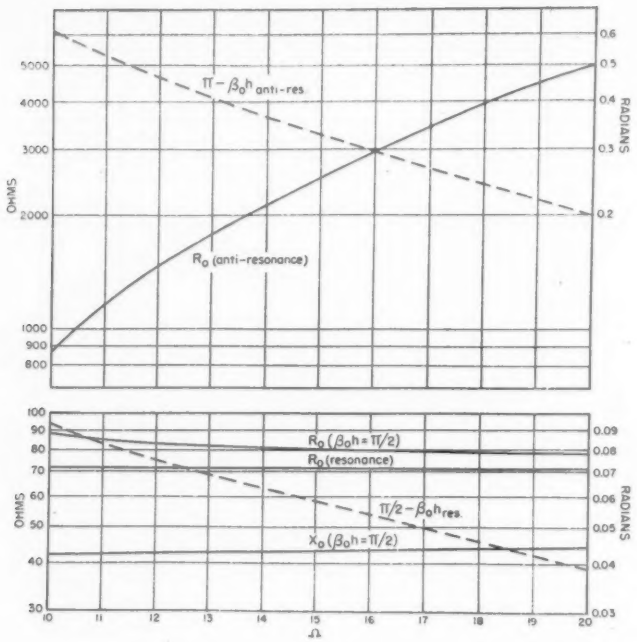


FIG. 13. Second order resonant and anti-resonant values of R_0; R_0 and X_0 at $\beta_0 h=\pi/2$; $\pi/2-\beta_0 h_{res.}$ and $\pi-\beta_0 h_{anti-res.}$ as functions of Ω.

The first and second order correction factors in the form $\delta\epsilon/h$ are shown in Figs. 3–6 as functions of $\beta_0 h$ for three values of $\Omega=2ln(2h/a)$. It is seen that the correction is significant primarily near $\beta_0 h=\pi/2$, $3\pi/2$, etc. Curves showing second order values of R_0-R_δ and X_0-X_δ are found in Figs. 7–9 for several values of $\beta_0\delta$ and with $\Omega=10$, 15, 20.

5. Conclusion. The King-Middleton theory for the cylindrical antenna has been generalized to show the effect of a finite gap on the current and the impedance. The analysis shows that for small gaps the impedance is not very sensitive to gap length so that impedances calculated for zero gap are good approximations for a large class of antennas for which $\beta_0\delta<0.01$. If $\beta_0\delta\geqq 0.01$ the correction curves of Figs. 7–9 may be used in conjunction with the curves of Figs. 10–13 for R_0 and X_0. It is interesting to note that for $\beta_0 h$ near $\pi/2$ the impedance Z_δ corrected for the gap differs negligibly from the uncorrected value Z_0 if the actual conductor half-length $h-\delta$ is used instead of h. This is not true for $\beta_0 h$ near π.

AERODYNAMICS OF THIN QUADRILATERAL WINGS AT SUPERSONIC SPEEDS*

BY

ROBERT M. SNOW

The Johns Hopkins University

1. Introduction. The method of conical fields, due to Busemann,[1] furnishes a powerful tool for the study of linearized supersonic flow around obstacles of special type. An appropriate application is to plane (infinitely thin) wings at small angles of attack. The present paper is devoted to a study of the supersonic aerodynamics of some of the more interesting special cases of quadrilateral planform. The application of the formulae obtained to a more general polygonal planform is suggested, and a procedure is given for accurately taking account of a dihedral bend in the wing. Modifications due to wing thickness, viscous effects, and interference effects (with a fuselage or with other wings) are beyond the scope of this method.

Since the general theory has been discussed by Stewart[2] recently in this journal, a brief description of the method will suffice. A conical field corresponds to a linear homogeneous solution of the linearized (or Prandtl-Glauert) potential equation for supersonic flow:

$$\frac{\partial^2\Phi}{\partial x^2}+\frac{\partial^2\Phi}{\partial y^2}-(M^2-1)\frac{\partial^2\Phi}{\partial z^2}=0. \tag{1}$$

Here M is the Mach number of the main stream, which is moving along the z-axis. The perturbation velocity components (u, v, w) are also solutions of Eq. (1) and are homogeneous of degree zero, i.e., u, v, and w are constant along any ray emanating from the origin. The particular simplicity of conical fields lies in the fact that after a transformation[3] the perturbation velocity components are obtained a ssolutions of Laplace's equation in two variables. Let x, y, z be rectangular coordinates with respect to a set of axes fixed in the wingtip, with the z-axis pointing downstream, the y-axis normal to the wing, and the x-axis directed spanwise away from the wing. If

$$\phi=\arctan y/x, \qquad R=\sqrt{x^2+y^2}/z, \qquad A=\tan\mu=(M^2-1)^{-1/2},$$

then the transformation

$$r=\left\{\frac{1-\sqrt{1-R^2/A^2}}{1+\sqrt{1-R^2/A^2}}\right\}^{1/2}=\frac{A}{R}-\sqrt{\frac{A^2}{R^2}-1}, \qquad R=2Ar/(1+r^2) \tag{2}$$

---

* Received Oct. 28, 1946. This paper is based on work done for the Bureau of Ordnance under contract NORD 7386. The author is indebted to Dr. L. L. Cronvich for valuable discussion and advice.

[1] A. Busemann, *Infinitesimale kegelige Ueberschallstroemung*, Schriften Deutsch. Akad. Luftfahrtforschung, **7**, 105–121 (1943).

[2] H. J. Stewart, *The lift of a Delta wing at supersonic speeds*, Quart. Appl. Math. **4**, 246–254 (1946).

[3] Busemann credits this transformation to Chaplygin, who made use of it in a formally similar problem. Cf. S. Chaplygin. *Gas jets*, Part V, Scientific Memoirs, Moscow University, 1902. Translated from the Russian as NACA Tech. Memo. No. 1063. See also the translation published in mimeographed form by Brown University in 1944.

is such that homogeneous functions of degree zero which satisfy Eq. (1) also satisfy Laplace's equation in the polar coordinates (r, ϕ). The evaluation of the streamwise component (w) of perturbation velocity is of primary importance since the aerodynamic forces on the wing are determined by w alone. This follows from the linearized Bernoulli equation,

$$p = p_0 - \rho_0 W w, \tag{3}$$

which like Eq. (1), results from neglecting squares and products of perturbation velocities in the corresponding exact equation, In many problems of this type, including those considered here, the boundary conditions are such that w may be determined directly without further reference to the components u and v.

2. **Characteristic cones. Boundary conditions.** To illustrate the type of boundary condition needed to determine the conical field, consider the special case of a rectangular wing. This rectangular wing may be regarded as the result of cutting the ends off of a two-dimensional airfoil. This operation causes a modification in the (originally two-dimensional) flow; this modification may be referred to as the "tip effect." In this connection a fundamental distinction should be made between subsonic flow and supersonic flow. For subsonic flow (differential equation of elliptic type) the tip effect dies off asymptotically with increasing distance inboard. For

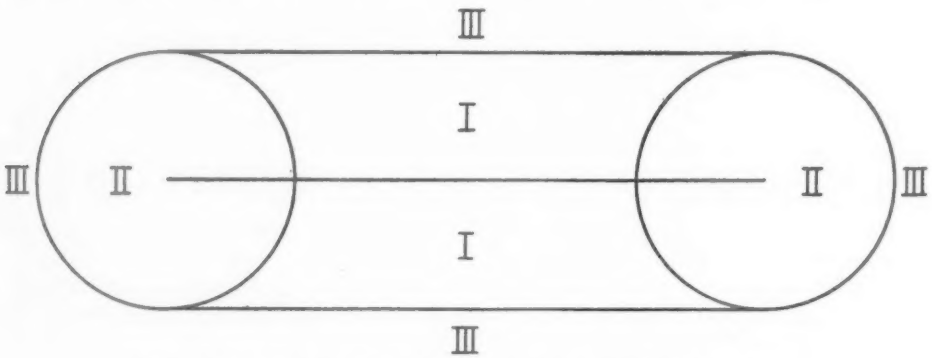


FIG. 1. Flow for rectangular wing, section by plane perpendicular to stream.
I—Flow same as for flow about infinite span airfoil.
II—Conical field.
III—Free stream.

supersonic flow (differential equation of hyperbolic type) the tip effect falls to zero at a certain finite distance, and the entire effect is contained within a region bounded by real characteristic surfaces. For linearized supersonic flow, the domain of influence of any point is bounded by a "Mach cone," which is (one nappe of) a cone opening downstream with semi-vertex angle equal to the Mach angle μ. Fig. 1 represents a section by a plane perpendicular to the main stream. The Mach cones from the tips of the leading edge divide this plane into three types of regions. In the central region (I) the flow is in all respects the same as if the wing were of infinite span, since no point of this central region lies in the domain of influence of any point removed in the mental process of obtaining the rectangular wing from an airfoil of infinite span. On

the other hand the perturbation velocity components are zero in the exterior region (III), since no point of this region lies in the domain of influence of any point on the rectangular wing. The requirement of continuity leads to boundary conditions which must be satisfied by the perturbation velocity components u, v, w on the boundary of each conical transition region (II). In the boundary between region (II) and (III), u, v, and w must vanish. On the boundary between regions (I) and (II), u, v, w take on the (constant, two-dimensional) values of region (I).

Similar remarks apply to the example of a swept-back leading edge (Fig. 2). The two lines forming the leading edge are of course finite in the actual case, but the effect of their finiteness, so to speak, will be confined by characteristic cones passing through the points at which the leading edge changes direction. The possibility of treating

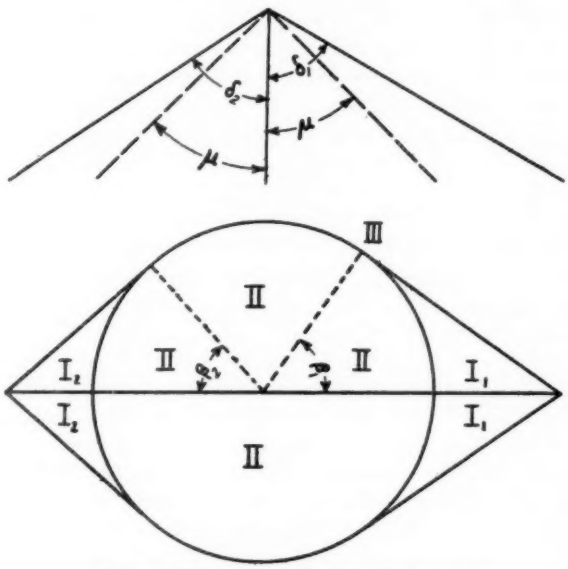


FIG. 2. Flow for swept-back leading edge.
Above, planform for definition of symbols. Airflow is toward bottom of page.
Below, section by plane perpendicular to stream.
I_1—Flow same as for infinite span airfoil at angle of yaw $\pi/2-\delta_1$.
I_2—Flow same as for infinite span airfoil at angle of yaw $\pi/2-\delta_2$.
II —Conical field.
III —Free stream.

more general polygonal wings by this method follows from these remarks. For the polygons which may be thus treated there are isolated regions of uniform flow (identical with the flow for an infinite wing at a certain angle of yaw) separated by regions of transition in the Mach cones which start from each vertex of the polygon and open downstream. It is convenient in the following to refer to these special Mach cones simply as "the Mach cones" or "the Mach cone."

Boundary conditions must also be given over that part of the wing which lies inside the Mach cone. The velocity component normal to the wing must have the

same value as on the rest of the wing, because of the condition that there be no flow through the wing. The boundary condition for w is that the normal derivative of w be zero on the wing. This follows from the requirement of irrotationality (implicit in the use of the potential Φ), so that $\partial w/\partial y = \partial v/\partial z$, and the condition that v is constant on the wing. The fact that the wing does not lie exactly in the plane $y=0$ is neglected; this has no effect on the first order perturbation. This simplification is made throughout, so that the angle of attack enters only in the boundary values and not in the position of the boundaries.

3. Bent leading edge. The basic flow problem which we shall solve is sketched in Fig. 2. The angles δ_1 and δ_2 are not necessarily acute, as in the case drawn, but each is assumed to lie in the range $\mu < \delta < \pi - \mu$. To simplify the immediate discussion it is assumed that the angle of the leading edge points upstream, so that $\delta_1 + \delta_2 < \pi$; it is shown in the next section that the formulae obtained are valid without this restriction. Evidently the wing cleaves the problem into two separate problems which may be treated independently. Attention may be confined to the upper half, since the solution for the lower half differs only in sign. The points on the circle for which $\phi = \beta_1$ and $\phi = \pi - \beta_2$ mark the tangency of the plane Mach waves from the leading edge with the Mach cone. By elementary geometry, $\cos \beta_1 = \tan \mu / \tan \delta_1$, $\cos \beta_2 = \tan \mu / \tan \delta_2$. From a consideration of the "domains of influence," it is seen that the flow in the regions I_1, I_2, and III is as indicated in the legend to Fig. 2. The normal derivative of w vanishes on the wing. All three components of perturbation velocity vanish on the arc $\beta_1 < \phi < \pi - \beta_2$; on the other two arcs the boundary conditions are to be obtained from the essentially two-dimensional problem of an infinite span airfoil at an angle of yaw, to which we now turn.

Let δ be the angle between the main stream direction and the leading edge of an airfoil of infinite span, so that $(\pi/2) - \delta$ is the angle of yaw. It is assumed that $\mu < \delta < \pi - \mu$. The uniform flow W may be considered as a superposition of a uniform flow at velocity $W \cos \delta$ parallel to the leading edge, giving rise to no perturbation, and a flow perpendicular to the leading edge at velocity $W_1 = W \sin \delta$ and effective Mach number $M_1 = M \sin \delta$. The effective angle of attack (α_1) for this second flow is measured in a plane perpendicular to the leading edge; it is related to the streamwise angle of attack (α), measured in a plane containing the stream direction and perpendicular to the plane of the wing for zero angle of attack, by the formula

$$\tan \alpha = \tan \alpha_1 \sin \delta.$$

Within the limits of validity of the linear theory we need not distinguish between the angle of attack and its tangent, so that $\alpha = \alpha_1 \sin \delta$. For a plane airfoil of infinite span and not yawed, the streamwise component (w) of perturbation velocity is

$$w = \alpha W \tan \mu = \alpha W / \sqrt{M^2 - 1} \equiv w_\infty.$$

To obtain the *chordwise* component of perturbation velocity for an airfoil at an angle of yaw $(\pi/2) = \delta$, replace α, W, M by α_1, W_1, M_1. The *streamwise* component of perturbation velocity follows from multiplication by $\sin \delta$:

$$w = \alpha_1 W_1 \sin \delta / \sqrt{M_1^2 - 1} = \alpha W \sin \delta / \sqrt{M^2 \sin^2 \delta - 1},$$

or

$$w = w_\infty/\sin\beta, \tag{4}$$

where, as before, $\cos\beta = \tan\mu/\tan\delta$.

To summarize, the boundary conditions for w on the upper semicircle are

$$\begin{aligned} w &= w_\infty/\sin\beta_1 \equiv K_1 \quad \text{for} \quad 0 < \phi < \beta_1, \qquad r = 1, \\ w &= 0 \qquad\qquad\qquad\quad \text{for} \quad \beta_1 < \phi < \pi - \beta_2, \; r = 1, \\ w &= w_\infty/\sin\beta_2 \equiv K_2 \quad \text{for} \quad \pi - \beta_2 < \phi < \pi, \; r = 1. \end{aligned}$$

The potential problem which is now uniquely determined in the upper semicircle may be written out immediately as a Fourier series. This is a cosine series only, because of the condition that $\partial w/\partial n = 0$ for $\phi = 0$ and for $\phi = \pi$.

$$\begin{aligned} w &= \frac{\beta_1 K_1 + \beta_2 K_2}{\pi} + \frac{2}{\pi}\sum_{n=1}^{\infty} r^n \cos n\phi \left\{ K_1 \int_0^{\beta_1} \cos nu\,du + K_2 \int_{\pi-\beta_2}^{\pi} \cos nu\,du \right\} \\ &= \frac{\beta_1 K_1 + \beta_2 K_2}{\pi} + \frac{K_1}{\pi}\left\{ \arctan \frac{r\sin(\phi+\beta_1)}{1 - r\cos(\phi+\beta_1)} - \arctan \frac{r\sin(\phi-\beta_1)}{1 - r\cos(\phi-\beta_1)} \right\} \\ &\quad - \frac{K_2}{\pi}\left\{ \arctan \frac{r\sin(\phi+\beta_2)}{1 + r\cos(\phi+\beta_2)} - \arctan \frac{r\sin(\phi-\beta_2)}{1 + r\cos(\phi-\beta_2)} \right\}. \end{aligned} \tag{5}$$

In Eq. (5) each inverse tangent is restricted to its principal values ($-\pi/2$ to $+\pi/2$).

In the symmetrical case $\delta_1 = \delta_2$, the expression for w on the wing ($\phi = 0, \pi$) simplifies to

$$w = \frac{2K}{\pi}\left\{ \beta + \arctan \frac{r^2 \sin 2\beta}{1 - r^2\cos 2\beta} \right\} = \frac{2}{\pi}\,\frac{w_\infty}{\sin\beta} \arctan \frac{\tan\beta}{\sqrt{1 - R^2/A^2}}. \tag{6}$$

Returning to the general case (δ_1 and δ_2 not necessarily equal) we seek the average value of w along the segment of a spanwise line (i.e., perpendicular to the stream) cut off by the Mach cone. It is necessary to evaluate integrals such as

$$\frac{1}{A}\int_0^A \arctan \frac{r\sin\gamma}{1 - r\cos\gamma}\,dR.$$

Since $R/A = 2r/(1+r^2)$, integration by parts leads to

$$\frac{1}{A}\int_0^A \arctan \frac{r\sin\gamma}{1 - r\cos\gamma}\,dR = (\pi\tan\gamma)/4 + (1 - \sec\gamma)(\pi - \gamma)/2, \qquad (0 < \gamma < 2\pi).$$

Using this result it is found that the average value sought is

$$\bar{w} = [K_1(1 + \tan\beta_1 - \sec\beta_1) + K_2(1 + \tan\beta_2 - \sec\beta_2)]/2. \tag{7}$$

4. Case of intersecting envelopes. For treating forward sweep or dihedral, it is necessary to discuss the plane waves from the leading edge in more detail. On one side there is a weak shock wave and on the other side a weak expansion wave; however, in the linear theory the distinction between shock waves and expansion waves dis-

appears.* Both are regarded merely as surfaces of discontinuity, which can occur only across the *envelope* of Mach cones with vertex on the leading edge. As the simplest example Fig. 3 shows a section perpendicular to the main stream for the case of a wing with dihedral and with the leading edge perpendicular to the main stream. The situation for a plane wing with forward sweep would differ only in that the trace LMN of the wing in Fig. 3 would be straight, and overlapping of the plane waves would occur over the bottom arc as well as the top.

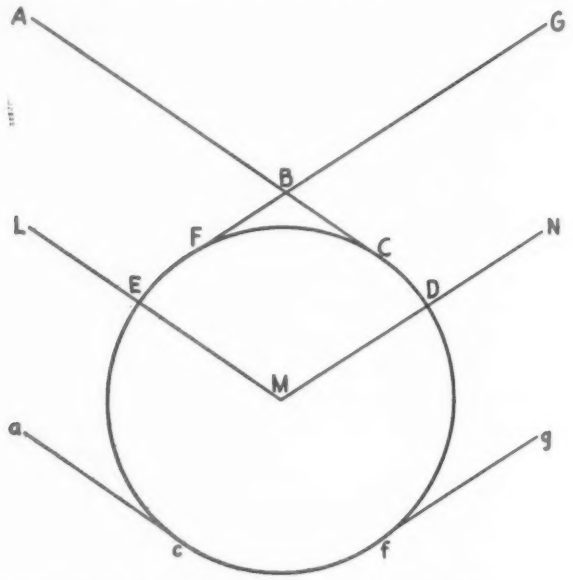


FIG. 3. Wing with dihedral, section by plane perpendicular to stream. In case drawn, leading edge is perpendicular to stream.

The envelope of all Mach cones with vertex on LM consists of two half planes with traces AC and ac. Similarly the Mach cones with vertex on MN give rise to the envelope represented by FG and fg. It is important to notice that no arc of the circle is part of either envelope. Within the bounds of the linear theory, shock waves or rarefaction waves intersect without mutual interference, and the perturbations caused by each are additive. In the region $GBCDN$ the flow is uniform; in the region $ABFEL$ there is another uniform flow; in the region between FBC and the circle the flow is also uniform, since the components of the perturbation velocity are obtained by addition of the components of perturbation velocity for the other two uniform flows. This completes the specification of the boundary conditions for the upper part of the circle. Whether we are dealing with u, v, or w, the sought function assumes a constant value of EF, another constant value of CD (both of these constants ob-

* The entropy increase is of the third order in the perturbation velocity components, whereas the linear theory retains only the first order. It is this fact which makes possible the usual two-dimensional linear and second order calculations, in which the pressure is determined by *local* conditions and does not depend on the history of the flow up to that point.

tainable from two dimensional theory), and a constant value of the arc FC, namely the sum of the other two constants. The boundary conditions for the lower part of Fig. 3 present nothing new; u, v, and w take on calculable (constant) values on the arcs Ec and fD, and the value zero on the arc cf.

Review of the problem illustrated by Fig. 2 now shows that the analysis given holds also for the case of a wing with forward sweep, i.e., with the angle pointing downstream; the only difference is a slight modification of Fig. 2.

5. Trapezoidal wing. We are now in a position to study the lift coefficient and center of pressure for the symmetrical trapezoidal wing shown in Fig. 4. The leading and trailing edges are perpendicular to the main stream, and the tip angle (δ) is

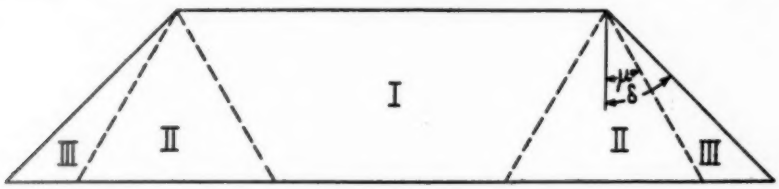


FIG. 4.

greater than the Mach angle. Since the leading edge of perpendicular to the stream, $\delta_2=\beta_2=\pi/2$, and the subscript may be dropped from δ_1 and β_1. In the region I, $\bar{w}/w_\infty=1$. In the region II, the average w is found from Eq. (7):

$$\bar{w}/w_\infty = 1/2 + (1 + \tan\beta - \sec\beta)/2\sin\beta.$$

In the region III, $\bar{w}/w_\infty=1/\sin\beta$. On taking the average of these quantities, weighted according to the area in which each applies, it is found that

$$C_L/C_{L\infty} = 1, \tag{8}$$

where $C_{L\infty}$ is the lift coefficient for an infinite span airfoil with leading edge perpendicular to the stream. Similarly the center of pressure is found to lie behind the leading edge by the distance

$$z = c(1 + c\tan\delta/3s)/2. \tag{9}$$

Thus in this case *the lift coefficient and center of pressure are the same as if the wing were subject to the uniform lift distribution of an infinite span airfoil.* The actual lift is not uniform; in the region I the lift is that of an infinite span airfoil; in the region II the

FIG. 5. Spanwise pressure distribution for wing shown in Fig. 4. $\delta=45°$, $\mu=30°$($M=2$).

life is less, and in the region III the lift is greater by just enough to compensate for the decreased lift in II. As an example, Fig. 5 shows the spanwise pressure distribution for the case $\delta=45°$, $\mu=30°$ $(M=2)$.

The above discussion, and the manner of drawing Figs. 4 and 5, assumed that the two Mach cones from the tips do not intersect on the wing. However, this restriction is seen to be unnecessary. If Φ_1 is the (disturbance) potential inside one of the cones, Φ_2 the potential inside the other cone, and Φ_∞ the potential for a wing of infinite span with leading edge perpendicular to the stream, then in the region common to the two cones the potential is $\Phi=\Phi_1+\Phi_2-\Phi_\infty$. To verify this, it may be noted that Φ_1, Φ_2 and Φ_∞ are solutions of the Prandtl-Glauert Eq. (1), and since that equation is linear, Φ is a solution as well. Also Φ and its first derivatives (u, v, w) are continuous across the conical surfaces bounding the region in question. We may say that the "tip effects" from the two tips are additive, since the equation defining Φ may be written

$$(\Phi_\infty - \Phi) = (\Phi_\infty - \Phi_1) + (\Phi_\infty - \Phi_2).$$

(It will be noticed that the flow in the region in question is not a conical field. In general, the result of superposing two conical fields with different vertices is not a conical field, although it approaches a conical field asymptotically downstream.) Utilizing this result, it is easily seen that the results for lift coefficient and center of pressure are unaffected by the overlapping of the two conical fields.

6. General symmetrical quadrilateral. We turn now to the problem of a quadrilateral which is symmetrical about a diagonal, that diagonal being parallel to the stream and of length c. The semi-vertex angles at the nose and tail, say δ and δ_1 respectively, are not necessarily acute angles (see Fig. 7 for the various possibilities); it is assumed only that each lies in the range μ to $\pi-\mu$. It is, of course, necessary that $\delta+\delta_1<\pi$.

The forward pointing triangle is a special case, $\delta_1=\pi/2$. It is also a special case of the trapezoid; setting $s_L=0$ or $s=c\tan\delta$ in Eqs. (8) and (9) leads to

$$C_L/C_{L\infty} = 1, \qquad z = 2c/3. \tag{10}$$

Here c is the distance from the vertex to the trailing edge. Fig. 6 shows the pressure distribution spanwise for the case $\delta=45°$, $\mu=30°$ $(M=2)$. Fig. 6 is, of course, applicable to any other case for which $\tan\delta=\sqrt{3}\tan\mu$, by a uniform change of scale.

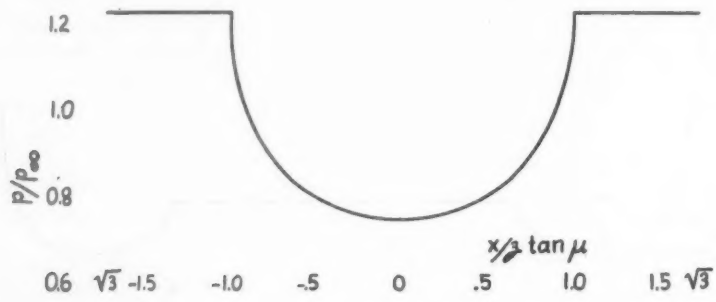


FIG. 6. Spanwise pressure distribution for forward pointing triangle. Semi-vertex angle $\delta=45°$. Mach angle $\mu=30°$ $(M=2)$.

For the general quadrilateral of Fig. 7, the pressure distribution is given by Eq. (6), in connection with Eq. (3). In deriving Eq. (6) it was perhaps tacitly implied that the planform of the wing consists of two *semi-infinite* lines, but a consideration of "domains of influence" shows that the requirement $\delta_1 > \mu$ is sufficient to ensure

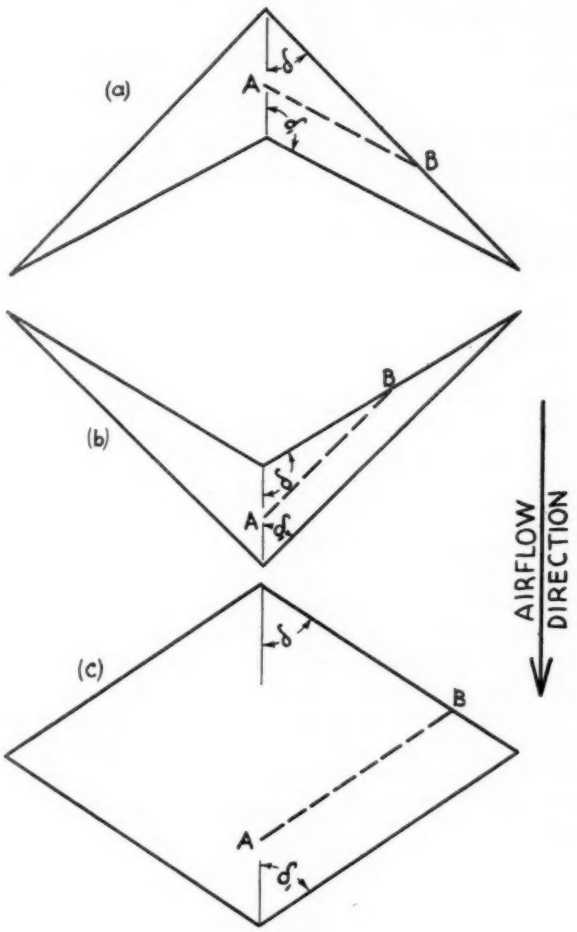


FIG. 7.

that Eq. (6) holds at every point on the wing. The most convenient procedure for finding the average pressure over the wing is to integrate along any line (AB in Fig. 7) parallel to the trailing edge on one side. This gives for the average value of w, by Eq. (6),

$$\bar{w} = \left\{\frac{w_\infty}{\sin\beta}\left[\int_0^{t_1} \frac{2}{\pi} \arctan\frac{\tan\beta}{\sqrt{1-\lambda^2}}\,dt + \int_{t_1}^{t_2} dt\right]\right\} \Big/ \int_0^{t_2} dt,$$

where

$$\lambda = R/A, \qquad t = \lambda \sec \delta_1/(\lambda + \sec \beta_1),$$

$$t_1 = t]_{\lambda=1} = \sin \mu/\sin (\mu + \delta_1), \qquad t_2 = t]_{\lambda=\sec \beta} = \sin \delta/\sin (\delta + \delta_1).$$

The integration leads eventually to

$$C_L/C_{L\infty} = \bar{w}/w_\infty = \frac{2}{\pi} \frac{\beta_1 \sin 2\beta - \beta \sin 2\beta_1}{\sin \beta_1 \sin 2\beta - \sin \beta \sin 2\beta_1}. \tag{11}$$

The angles β and β_1 are, of course, to be measured in radians. This expression for $C_L/C_{L\infty}$ is symmetric in the two angles β and β_1, and therefore *is unchanged by interchanging* δ *and* δ_1.

The center of pressure is at $z=2c(1-\bar{t} \cos \delta_1)/3$, where $\bar{t}$ is the weighted average of t along AB, with weights proportional to the pressure. It is found that

$$\bar{t} \cos \delta_1 = \frac{\displaystyle\int_0^{t_1} t \cdot \frac{2}{\pi} \arctan \frac{\tan \beta}{\sqrt{1-\lambda^2}}\, dt + \int_{t_1}^{t_2} t\, dt}{\displaystyle\int_0^{t_1} \frac{2}{\pi} \arctan \frac{\tan \beta}{\sqrt{1-\lambda^2}}\, dt + \int_{t_1}^{t_2} dt} \cos \delta_1 \tag{12}$$

$$= \frac{1}{2} \frac{\cos^2 \beta_1 + \cos^2 \beta}{\cos^2 \beta_1 - \cos^2 \beta} + \frac{\sin 2\beta}{2 \sin^2 \beta_1} \cdot \frac{\sin 2\beta_1 - 2\beta_1 \cos 2\beta_1}{2\beta_1 \sin 2\beta - 2\beta \sin 2\beta_1}. \tag{13}$$

These formulae contain as special cases the forward pointing triangle ($\beta_1=\pi/2$) and the backward pointing triangle ($\beta=\pi/2$). For these triangles, the invariance of lift coefficient (and in this particular case, the center of pressure also) with respect to reversal of the flow direction may be easily verified. It has already been shown that lift coefficient and center of pressure for the forward pointing triangle are the same as if the pressure were uniform (which it is not; cf. Fig. 6). For the backward pointing triangle the pressure *is* uniform. Reversal of direction of flow thus causes a radical change in pressure distribution in general, even though it does not alter the lift coefficient, for these quadrilateral wings.

Another special case of interest is the diamond ($\delta_1=\delta$, $\beta_1=\beta$). In this case Eqs. (11) and (12) reduce to

$$C_L/C_{L\infty} = (\sin 2\beta - 2\beta \cos 2\beta)/\pi \sin^3 \beta \tag{14}$$

$$\frac{z}{c} = \frac{1 - 2\beta \sin^2 2\beta/3(\sin 2\beta - 2\beta \cos 2\beta)}{1 - \cos 2\beta}. \tag{15}$$

Using Eqs. (14) and (15), the lift coefficient and center of pressure for a diamond have been evaluated and are presented in the following table. It is seen at once that the property of invariance under reversal of flow direction, which was found to hold for C_L, does not in general apply to the center of pressure. If such were the case the center of pressure of the diamond would necessarily be at $z=c/2$. From Table 1 it is seen that the center of pressure of a diamond actually lies forward of the midpoint, though never forward of $7c/15$ as long as $\delta > \mu$.

TABLE 1

β	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
C_L/C_{L_∞}	.8488	.8511	.8592	.8720	.8897	.9120	.9376	.9646	.9885	1.0000
z/c	.4667	.4671	.4658	.4709	.4743	.4788	.4842	.4905	.4966	.5000
$\tan\delta/\tan\mu$	1.000	1.015	1.064	1.155	1.305	1.556	2.000	2.924	5.759	∞

7. Dihedral. The case of a wing with dihedral, which may be combined with forward or backward sweep at the dihedral point, is reducible by a conformal transformation to the case of a simple bend in the leading edge of a flat wing. In Fig. 8, the angle $\phi=0$ has been taken in the wing (with no loss of generality). Let γ be the radian measure of the arc of the circle subtended by the wing. The part of the circle not drawn refers to an independent problem of the same type, with a different γ. The angles β_1 and β_2 have the same meaning as in the plane case. In the case shown, $\beta_1+\beta_2<\gamma$ (which incidentally corresponds to considerable sweepback in this case), but this condition is not essential, and is assumed here merely to simplify the drawing.

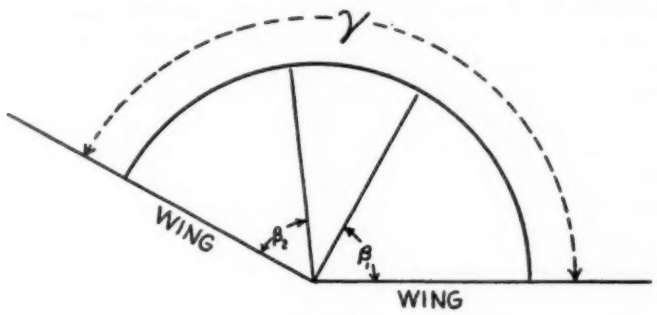


FIG. 8.

The potential problem to be solved in the r, ϕ plane involves the following boundary conditions:

$$\begin{aligned} \partial w/\partial\phi &= 0 \quad \text{for} \quad \phi = 0 \text{ or } \gamma, \ (0 < r < 1), \\ w &= K_1 \quad \text{for} \quad 0 < \phi < \beta_1, \qquad r = 1, \\ w &= 0 \quad \text{for} \quad \beta_1 < \phi < \gamma - \beta_2, \ r = 1, \\ w &= K_2 \quad \text{for} \quad \gamma - \beta_2 < \phi < \gamma, \ r = 1. \end{aligned}$$

The transformation $\epsilon' = \epsilon^{\pi/\gamma}$ maps the relevant part of the unit circle in the $\epsilon = re^{i\phi}$ plane into the upper half of the unit circle in the $\epsilon' = r'e^{i\phi'}$ plane. It is clear that w is given by Eq. (5) with r replaced by $r^{\pi/\gamma}$, and every angle (β_1, β_2, ϕ) multiplied by π/γ.

It remains to specify $\beta_1, \beta_2, K_1, K_2$ in terms of the geometry of the wing. As in the plane case,

$$\cos\beta_1 = \tan\mu/\tan\delta_1, \qquad \cos\beta_2 = \tan\mu/\tan\delta_2,$$

where δ_1 and δ_2 are the angles from the main stream direction to the leading edge. The

boundary values K_1 and K_2 are given by the same expression as for a plane wing, except that the angle of attack is now, in general, different for each plane of the wing. Writing α_1 and α_2 for these "local" angles of attack, one finds that

$$K_1 = \alpha_1 W \tan \mu / \sin \beta_1, \qquad K_2 = \alpha_2 W \tan \mu / \sin \beta_2.$$

It is interesting to note that in the case of symmetry ($\delta_1 = \delta_2$, $\alpha_1 = \alpha_2$) there is a certain dihedral, namely $\gamma = 2\beta$, for which the boundary condition is $w = K$ over the entire arc γ so that the flow is uniform in the whole sector.

The restriction $\beta_1 + \beta_2 < \gamma$ is seen to be non-essential as in the previous case of a bent leading edge. Also it should be pointed out that for a wing with upswept dihedral ($\gamma < \pi$ for the upper surface) the lift is decreased in magnitude.

These formulae may be applied to a perpendicular vane at the tip of a rectangular wing, the vane being large enough to project through the Mach cone. This may be regarded as an example of a wing with dihedral, one of the angles of attack being zero.

If the vane extends both above and below the wing, the lift remains constant to the end of the wing. If the vane is confined to either the top or the bottom of the wing, the lift decreases moving along a span line toward the wingtip; the lift at the tip is 1/3 of the lift in the central region of the wing, and the spanwise average from tip to Mach cone is found by integration to be $(1 - 2/3\sqrt{3})$ of the lift in the central region.

ON AXIALLY SYMMETRIC FLOWS*

BY

ALEXANDER WEINSTEIN

Carnegie Institute of Technology

1. Introduction. The determination of the irrotational flow of a perfect incompressible fluid around a given body constitutes a boundary value problem which can be solved by methods of Potential Theory. However, this theoretical solution has found little practical application.

In the case of a solid of revolution, an indirect but more efficient approach is given by the *method of sources and sinks*, which may be called an inverse method in the same sense in which this term is used in Elasticity and elsewhere in Hydrodynamics. The body of revolution cannot be prescribed but only approximated to a certain extent by the flow due to a suitable distribution of sources and sinks in joint action with a parallel uniform flow coming from infinity. As a compensation for this drawback, the approximating flows are given by explicit formulae which are often suitable for numerical computations.

The method of sources and sinks was originally confined to the case of plane motion until Stokes' generalization of the concept of stream function enabled Rankine in 1871 to adapt the method to axially symmetric flows. In the succeeding decades various examples were given and applied to the pressure distribution around airships. However, the possibilities implied by the method were far from being exhausted. In fact, up to the present day no other flows have been considered but such as are due to a distribution of sources, sinks, and doublets exclusively on the axis of symmetry. Let us consider, for instance, the case in which the direction of the parallel flow coincides with the axis of symmetry. In this case—the only one which will be discussed in the present paper—it has been already noticed by Munk[1] that blunt nosed bodies cannot be obtained by taking any distribution of sources and sinks on the axis.

The present paper deals with an *extension of the method of sources and sinks*. The sources and sinks are no longer confined to the axis but are distributed on circumferences, rings, discs and cylinders. The distribution must of course be symmetric with respect to the axis of revolution, but for practical purposes the choice is restricted to such cases in which the stream function can be explicitly computed in terms of known functions. We shall use here Beltrami's fundamental results obtained in a series of papers published in 1878–80. The importance of Beltrami's results for the theory of hydrodynamical flows has been completely overlooked. A serious error in his paper requiring a modification of nearly all his formulae does not seem to have been noticed. In fact, Beltrami, who applies his formulae chiefly to problems of Potential Theory and Electrostatics, fails to recognize that Stokes' stream function is, in many of his formulae, a many valued function. It is interesting to note that this

* Received Dec. 12, 1946. This paper was written in connection with work done at Harvard University at the request of the Bureau of Ordnance, U. S. Navy. The author wishes to express his gratitude to Professor Garrett Birkhoff for his kind interest and cooperation.

[1] M. M. Munk in *Aerodynamic theory* (edited by W. F. Durand) vol. 1, J. Springer, Berlin, 1934, p. 266.

property of Stokes' function, which should be expected to hold rather as a rule than as an exception, is never mentioned in the literature on Hydrodynamics while the corresponding property of the stream function for plane motion is one of the basic concepts of the theory.

A superposition of the flow due to our sources and of a uniform flow in the direction of the axis of symmetry will give us some essentially new types of flows, in particular, flows around blunt nosed profiles. Up to the present date, only isolated limiting cases of these flows have been discussed. Incidentally, the use of sources and sinks outside the axis will enable us to obtain profiles consisting of piecewise analytic curves. As is well known, the profiles corresponding to a source distribution along the axis consist necessarily of a single analytic curve, the nose and the tail being the only possible singular points.

2. Stokes' stream function. In this paper we consider only steady irrotational axially symmetric flows of a perfect incompressible fluid. The axis of symmetry will be taken as the x-axis. Let x, ρ be the coordinates in a meridian plane. The flow is completely determined if the velocity distribution is known in the half plane $-\infty < x < +\infty$, $\rho \geqq 0$. The motion being irrotational, there exists a velocity potential ϕ which is an harmonic function satisfying the Laplace equation in three-space. For axially symmetric flows, ϕ is a function only of x and ρ, so that the Laplace equation in cylindrical coordinates for $\phi(x, \rho)$ reduces to

$$\frac{\partial}{\partial x}\left(\rho \frac{\partial \phi}{\partial x}\right) + \frac{\partial}{\partial \rho}\left(\rho \frac{\partial \phi}{\partial \rho}\right) = 0. \tag{2.1}$$

This equation can be interpreted as an integrability condition showing the existence of a function $\psi(x, \rho)$ which is called the *Stokes stream function* and which is defined by the equations

$$\frac{\partial \psi}{\partial x} = -\rho \frac{\partial \phi}{\partial \rho}; \qquad \frac{\partial \psi}{\partial \rho} = \rho \frac{\partial \phi}{\partial x}. \tag{2.2}$$

It is obvious, by (2.1), that $d\psi = (\partial\psi/\partial x)dx + (\partial\psi/\partial\rho)d\rho$ is a total differential and that the elimination of ϕ in (2.2) gives for $\psi(x, \rho)$ the differential equation

$$\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \psi}{\partial x}\right) + \frac{\partial}{\partial \rho}\left(\frac{1}{\rho}\frac{\partial \psi}{\partial \rho}\right) = 0. \tag{2.3}$$

Unlike ϕ, which is defined in three-space, the stream function $\psi(x, \rho)$ is defined only in the upper half plane $\rho \geqq 0$ by its total differential, i.e., by the differential equations (2.2). It is therefore to be expected that ψ will as a rule be a many-valued function. The hydrodynamical significance of the stream function ψ is well known: ψ remains constant along each streamline in the meridian plane, while $2\pi\psi$ represents the flow of a fluid of unit density between the given streamline and the streamline $\psi = 0$.

The velocity components u and v in the x and ρ direction are given by

$$\begin{aligned} u &= \frac{\partial \phi}{\partial x} = \frac{1}{\rho}\frac{\partial \psi}{\partial \rho}, \\ v &= \frac{\partial \phi}{\partial \rho} = -\frac{1}{\rho}\frac{\partial \psi}{\partial x}. \end{aligned} \tag{2.4}$$

By the convention adopted in this paper, the velocity vector is the gradient of the potential.

For a uniform flow parallel to the x-axis we have

$$\phi = Ux, \qquad \psi = \tfrac{1}{2}U\rho^2. \tag{2.5}$$

The flow $2\pi\psi$ is given by $\pi\rho^2 U$.

Beltrami calls ψ the associated function of ϕ. He shows, by a process of repeated differentiation and integration, that every potential or associated function generates a descending and an ascending sequence of pairs of functions of the same kind. This discovery is an important extension of the ordinary theory of analytic functions, and can be applied in a similar way to other cases.[2] However, our main interest will be focused on another part of Beltrami's work, in which he investigates fundamental solutions of (2.1) possessing basic singularities. The role played by such solutions in the theory of partial differential equations is too well known to be emphasized here, and was clearly recognized by Beltrami.

3. Sources and sinks on the axis. By our definition of the potential, the elementary potential in three-space is $-1/r$. In cylindrical coordinates x, ρ the potential ϕ of a source of strength m at the origin is given by

$$\phi_0 = \frac{-m}{\sqrt{x^2+\rho^2}}. \tag{3.1}$$

The corresponding stream function ψ is given by

$$\psi_0 = -m\left(1 + \frac{x}{\sqrt{x^2+\rho^2}}\right). \tag{3.2}$$

The additive constant is chosen in such a way that ψ_0 vanishes on the streamline $x<0$, $\rho=0$. We note that the stream function is single-valued in the upper half plane x, ρ.

4. The principle of superposition. The potential and the stream function of two or several combined flows is the sum of the respective functions corresponding to the separate flows. The method of sources and sinks uses combinations of flows due to sources and sinks and of a uniform parallel flow. A suitable choice leads to a flow with a streamline $\psi=0$ separating the domain of regular flow from the sources. This *dividing line* plays the part of a *rigid profile*, the interior of which can be replaced by a solid body. The principle of superposition for potentials is less important: the method of sources and sinks being based on the existence of the stream function. Closed dividing lines can be obtained only by taking a distribution of sources and sinks of total strength zero. For a positive total strength, the corresponding profile is open downstream and extends to infinity. Nevertheless, the corresponding surfaces of revolution, called *half-bodies*, are of great importance because in reality the potential flow is disturbed by viscosity everywhere except in a certain vicinity of the nose, which is at the same time the region in which big velocities and a danger of cavitation may be expected. The simplest case of a half-body will be discussed in the next paragraph.

5. Blasius-Fuhrmann's half-body. By superposition of a source m at the origin

[2] L. Bers and A. Gelbart, Trans. Amer. Math. Soc., **56**, 67–93 (1944).

and of a uniform stream U in the x direction, we obtain a flow with the potential and stream functions:

$$\phi = Ux - \frac{m}{\sqrt{x^2 + \rho^2}} = Ux - (m/r), \tag{5.1}$$

$$\psi = \tfrac{1}{2}U\rho^2 - m\left(1 + \frac{x}{\sqrt{x^2 + \rho^2}}\right) = \tfrac{1}{2}U\rho^2 - m(1 + \cos\theta), \tag{5.2}$$

where r and θ denote polar coordinates in the meridian plane. There will be a stagnation point

$$x_N = -\sqrt{\frac{m}{U}} \tag{5.3}$$

on the negative x-axis, which is obtained by putting $u(x, 0)$ equal to zero. We see from (5.2) that the streamline $\psi = 0$ consists of the negative x-axis and of the dividing line

$$\rho^2 = \frac{2m}{U}(1 + \cos\theta), \qquad \left(\frac{m}{U} = x_N^2\right), \tag{5.4}$$

which contains the stagnation point x_N. For $x = +\infty$ we have $\theta = 0$ and $\rho = \rho_\infty = 2\sqrt{m/U}$. The radius ρ_0 of the main parallel, as we shall call it, is obtained by putting $x = 0$ and $\theta = \pi/2$. We find

$$\rho_0 = \sqrt{\frac{2m}{U}} = \sqrt{2}\,|\,x_N\,| = \frac{1}{\sqrt{2}}\rho_\infty. \tag{5.5}$$

A somewhat laborious computation shows that the curvature K of the profile is steadily decreasing from the value $(9U/16m)^{1/2}$ at the stagnation point to zero at infinity.

6. Distributed sources along the axis. Various shapes of bodies and half-bodies have been obtained by Fuhrmann,[3] Kármán[4] and their followers by using continuous distributions of sources and sinks along the x axis.

Denoting by $q(x)$ the density of the source distribution of an interval $0 \leqq x \leqq l$ and by U the velocity of the parallel flow and using (3.2), we obtain the following expression for the stream function ψ of the resulting flow:

$$\psi(x, \rho) = \tfrac{1}{2}U\rho^2 - \int_0^l q(\xi)\left\{1 + \frac{x - \xi}{\sqrt{(x - \xi)^2 + \rho^2}}\right\}d\xi. \tag{6.1}$$

The integral $\int_0^l q(\xi)d\xi$ represents the total strength m of all sources and sinks and is zero in the case of a closed profile. Point sources and sinks may be included in the formula without difficulty. The profile is given by the equation $\psi = 0$. By taking a constant positive density $q(x) = m/l$, we obtain a half-body corresponding to the stream function

$$\psi_l = \tfrac{1}{2}U\rho^2 - m\left(1 + \frac{r_0 + r_l}{l}\right), \tag{6.2}$$

where r_0 and r_l denote the distances of (x, ρ) from $x = 0$ and $x = l$.

[3] G. Fuhrmann, Dissertation, Goettingen, 1912.

[4] Th. von Kármán, Abhandlungen aus dem aerodynamischen Institut, Aachen, 1927.

Fuhrmann computed several profiles by taking for $q(x)$ a step function or a piecewise linear function of x. He used also some additional point sources. It is obvious that the influence of sources on the axis is predominantly felt only on nearby portions of the profile.

For a *given* profile an approximation can be obtained, according to Kármán, by subdividing an interval $0 \leqq x \leqq l$ in the interior of the profile into a finite number n of not necessarily adjacent subintervals. Let l_k denote the length of the kth subinterval and let $q_k = m/l_k$ be the constant density of sources assigned to this segment. Let us choose n points $x_1, \rho_1, \cdots, x_n, \rho_n$ on the given profile. In order that the profile $\psi = 0$ given by our sources and the parallel flow shall pass through the n prescribed points, we must have [see (6.2)]

$$\sum_{k=1}^{n} m_k \left(1 + \frac{r_{ik}' - r_{ik}''}{l_k}\right) = \tfrac{1}{2} U \rho_i^2; \qquad i = 1, 2, \cdots, n, \tag{6.3}$$

where r' and r'' denote the distance from the endpoints of the kth subinterval to the point x_i, ρ_i on the profile. In this way, n linear non-homogeneous equations for $m_1, \cdots, m_n$ are obtained, which can be solved provided that the determinant of the coefficients

$$c_{ik} = 1 + \frac{r_{ik}' - r_{ik}''}{l_k}$$

is different from zero. In practical cases all l_k are taken equal in order to reduce the already very considerable computational work.

For a given profile, the ordinate $\rho = \rho(x)$ is a given function of x. Substituting this function in (6.1) and putting $\psi[x, \rho(x)] = 0$, we obtain an integral equation for the determination of the unknown density $q(x)$. Unfortunately, the resulting integral equation is a Fredholm equation of the first kind and, up to now, has proved to be useless. In fact, it is nearly obvious that a distribution along the axis can give only a limited number of different types of profiles.

The restriction imposed by the exclusive use of sources and sinks distributed on the axis has prevented so far any further development of the method. In the next paragraphs we shall remove this restriction by considering symmetrical distributions located outside of the x-axis.

7. The potential of a homogeneous circumference. As already mentioned in the Introduction our investigations will be based to a certain extent on Beltrami's results which we shall present here with the revisions, corrections and extensions required for our purposes.

Let us consider an axially symmetric distribution of sources and sinks, not necessarily located on the x-axis. It is clear that in the present case the potential ϕ_c of a homogeneous circumference with its axis on the x-axis will play the same role as the potential of a point source plays in the general case of an arbitrary three dimensional flow. Beltrami[5] gives two expressions for this elementary potential with axial symmetry about the axis. For our purposes we shall use the expression involving Bessel functions as more suitable for computation of the associated stream function.

[5] E. Beltrami, *Opere matematiche*, vol. 3, U. Hoepli, Milano, 1911 (especially pp. 349–382).

The alternative formula for ϕ_c and its relation to Laplace's expression for a potential has been recently investigated by Bateman and Rice.[6]

Beltrami's original proof of his formula for ϕ_c is complicated and will be replaced here by a more elementary proof covering several similar cases. This proof, which in its original form, was criticized by Watson[7] will be presented here with a slight addition which makes the reasoning convincing. (See also H. Bateman[8] who apparently accepts Watson's criticism.)

Bessel's Functions and Potential Theory. It is well known that the elementary method of separation of variables shows immediately that $e^{\pm zs}\ J_0(\rho s)$ is, for every value of the real parameter s, a solution of the equation (2.1) for a symmetric potential. ($J_0(z)$ denotes the Bessel function of index 0.) Throughout this paper we shall use Watson's notations. The function $J_0(z)$ is regular for every value of z and takes the value 1 for $z=0$.

Consider the potential function represented by the definite integral

$$\int_0^\infty e^{-|x|s} J_0(\rho s) F(s) ds, \qquad (x \neq 0) \tag{7.1}$$

in which the function $F(s)$ is supposed to be one which ensures uniform convergence and makes the limit of (7.1), as ρ tends to zero, equal to the result of making $\rho=0$ under the integral sign. When $x \neq 0$, this function takes the value

$$f(x) = \int_0^\infty e^{-|x|s} F(s) ds \tag{7.2}$$

on the x axis and may often be identified from the form of $f(x)$. In fact, we know from the elements of the theory of developments of potential functions in series, that an axially symmetric harmonic function is *uniquely* determined by the values it takes on a segment of the axis of symmetry.

Taking $F(s)=1$, we have $f(x)=|x|^{-1}$ by (7.2). The corresponding harmonic function is r^{-1}. By (7.1) we have, therefore, following formula

$$\int_0^\infty e^{-|x|s} J_0(\rho s) ds = \frac{1}{\sqrt{x^2+\rho^2}} \tag{7.3}$$

which will be used immediately. This formula is due to Lipschitz (Watson,[7] p. 384).

Let us now consider a homogeneous circumference C of radius b and of unit density with its center at the origin and its axis coinciding with the x-axis. The total mass (or total strength of sources) of C is $M=2\pi b$. By adding the elementary potentials due to the elements of C, we see at once that the potential of C at a point x of the axis is equal to

$$f(x) = -\frac{2\pi b}{\sqrt{x^2+b^2}}, \tag{7.4}$$

where $\sqrt{x^2+b^2}$ is the distance of x from any point of C. By using (7.2), (7.3), and (7.1) we immediately obtain the following expression for the potential ϕ_c of C:

[6] H. Bateman and S. O. Rice, American Journal of Mathematics, **60**, 297–308 (1938).

[7] G. N. Watson, *Theory of Bessel functions*, Cambridge University Press, 1922, p. 388.

[8] H. Bateman, *Partial differential equations*, Cambridge University Press, 1932, §7.32.

$$\phi_c(x, \rho) = -2\pi b \int_0^\infty e^{-|x|s} J_0(\rho s) J_0(bs) ds. \tag{7.5}$$

8. The stream function for a homogeneous circumference. In order to obtain the stream function ψ_c for our circumference C, let us consider the function

$$\psi_c^- = -2\pi b\rho \int_0^\infty e^{-|x|s} J_1(\rho s) J_0(bs) ds \tag{8.1}$$

for $x \leqq 0$. The Bessel function $J_1(z)$ vanishes for $z=0$ and satisfies the equation $J_1(z) = -J_0'(z)$. It is easily checked by (8.1) and (2.2) that ψ_c^- is the associated function to the potential ϕ_c as defined by (7.5), and it is obvious that $\psi_c^-(x, 0)$ vanishes for $x \leqq 0$. Similarly,

$$\psi_c^+(x, \rho) = 2\pi b\rho \int_0^\infty e^{-|x|s} J_1(\rho s) J_0(bs) ds \tag{8.2}$$

is for $x \geqq 0$, an associated function of (7.5). However, ψ_c^- and ψ_c^+ do not coincide on the positive ρ axis. In fact, we have as a special case of the discontinuous integral of Weber and Schafheitlin (Watson,[7] p. 406)

$$\int_0^\infty J_1(\rho s) J_0(bs) ds = \begin{cases} \dfrac{1}{\rho} & \text{for} \quad \rho > b \geqq 0, \\ 0 & \text{for} \quad b > \rho \geqq 0. \end{cases} \tag{8.3}$$

It follows that

$$\psi_c^-(-0, \rho) = \begin{cases} -2\pi b & \text{for} \quad \rho > b \geqq 0, \\ 0 & \text{for} \quad b > \rho \geqq 0, \end{cases} \tag{8.4}$$

and

$$\psi_c^+(+0, \rho) = \begin{cases} 2\pi b & \text{for} \quad \rho > b \geqq 0, \\ 0 & \text{for} \quad b > \rho \geqq 0. \end{cases} \tag{8.5}$$

Let us consider a simple closed curve L in the half plane $\rho \geqq 0$ containing the critical point $x=0$, $\rho=b$ (i.e., the trace of the circumference C) in its interior. The function ψ_c^+ is obviously the analytic continuation of ψ_c^- across the segment $0 \leqq \rho < b$. However, if we describe L counterclockwise and come back to a point $+0, \rho$, with $\rho > b$, we obtain a final value which is by $4\pi b$ greater than the initial value of ψ_c^-. It follows that *the stream function for a homogeneous circumference C is a many-valued function in the half-plane $\rho \geqq 0$ with the period $4\pi b$.* Since $2\pi b$ is the total strength M of the distributed sources, we can say that *the period is equal to $2M$.*

By rotation about the x axis the closed line L generates an axially symmetric torus, enclosing the sources on C. By the definition of the stream function, the outward flow across this torus is equal to 2π times the period of the stream function, i.e., this flow is equal to $4\pi M$, in perfect agreement with Gauss' theorem.

A single valued branch of the stream function ψ_c can be obtained in the domain D_b bounded by the x-axis and by the segment $x=0$, $0 \leqq \rho \leqq b$ by putting

$$\begin{aligned} \psi_c &= \psi_c^- \quad \text{for} \quad x \leqq 0, \\ \psi_c &= \psi_c^+ - 4\pi b \quad \text{for} \quad x \geqq 0. \end{aligned} \tag{8.6}$$

This function is continuous in D_b, vanishes on the negative x-axis and, by (8.4), takes the constant value $-2\pi b$ on the streamline $x=0$, $\rho > b$. For a circumference of total strength m and of uniform density $m/2\pi b$, we would have

$$\psi_c(0, \rho) = -m, \quad \text{for} \quad \rho > b. \tag{8.7}$$

For $\rho > b$ the constant value of $\psi_c(0, \rho)$ is equal to $-m$, independently of the radius b of C.

Comparing our result with Beltrami's formulae for the same stream function, we see that on p. 355 of the paper quoted in Footnote 5 Beltrami, failing to notice that ψ_c is a many-valued function, puts the wrong branches together and obtains a discontinuity along the ρ-axis. The same error occurs in A. G. Webster.[9] The formula given by Bateman[8] (p. 417; example 1) is also inaccurate.

The velocity u on the x axis due to the sources on C can be easily obtained by differentiating the potential (7.4) with respect to x. Denoting by m the total strength of C, we find

$$u(x, 0) = \frac{mx}{(x^2 + b^2)^{3/2}}, \tag{8.8}$$

so that u is zero for $x=0$. The maximum of $|u|$ is attained for $x^2 = \frac{1}{2}b^2$, where $|u| = 2 \cdot {}^{-2/3} m b^{-1/3}$.

9. The discontinuous integral of Weber-Schafheitlin. We give in this paragraph a new and simple proof for the evaluation of the integral of Weber-Schafheitlin [see (8.3)]. This proof is based on the general principles of Potential Theory and does not require any extensive technical knowledge of Bessel function.

As in Sec. 8 let us consider a homogeneous circumference C of radius b and unit total strength, so that $2\pi b = 1$. Let S be a sphere with center at the origin and with radius $R > b$. By Gauss' fundamental theorem, the outward flow across S is 4π. The surface of S cuts the half-plane $\rho > 0$ in a half-circle H. For $x \leqq 0$, the stream function $\psi_c(x, \rho)$ is given by (8.1). It takes the value zero along the negative x axis and up to the factor 2π, its value for a point P of H is given by the inward flow passing through the spherical cap generated by the rotation of the arc P, $-R$. The flow across a hemisphere being -2π, we have $\psi_c(0, R) = -1$. We see by (8.1) that

$$\lim_{x=0} \int_0^\infty e^{-|x|s} J_1(Rs) J_0(bs) ds = R^{-1}, \quad \text{for} \quad R > b. \tag{9.1}$$

On the other hand, it is known [and used in the proofs of (8.3)] that the left hand side of (8.3) can be obtained by putting $x=0$ in the integral (9.1). In this way we obtain

$$\int_0^\infty J_1(Rs) J_0(bs) ds = R^{-1}, \quad \text{for} \quad R > b.$$

In order to obtain the second part of (8.3) we have only to take a sphere with radius $R < b$.

[9] A. G. Webster, *Partial differential equations*, New York, Hafner Publishing Co., 1947 p. 368 ff.

10. The potential and the stream function for a disc. Let us consider a disc of sources of radius b with its center at the origin and its axis coinciding with the x-axis. Let $q(\rho)$ denote the areal density of the sources; the total strength m being given by

$$m = 2\pi \int_0^b \rho q(\rho) d\rho. \tag{10.1}$$

By the principle of superposition, the potential ϕ_d of the disc is obtained by integration of (7.5). Replacing b in this formula by a variable of integration η, we find, after multiplication of $q(\eta)d\eta$ and integration from 0 to b,

$$\phi_d(x, \rho) = -2\pi \int_0^b \eta q(\eta) d\eta \int_0^\infty e^{-|x|s} J_0(\rho s) J_0(\eta s) ds. \tag{10.2}$$

Putting

$$\chi(s) = \int_0^b \eta q(\eta) J_0(\eta s) d\eta \tag{10.3}$$

and interchanging the order of integration in (10.2), we obtain

$$\phi_d(x, \rho) = -2\pi \int_0^\infty e^{-|x|s} \chi(s) J_0(\rho s) ds. \tag{10.4}$$

Putting $\rho = 0$ and using Lipschitz' integral (7.3), we have

$$\phi_d(x, 0) = -2\pi \int_0^\infty e^{-|x|s} \chi(s) ds = -2\pi \int_0^b \frac{\eta q(\eta) d\eta}{\sqrt{x^2 + \eta^2}} \tag{10.5}$$

on the x axis, a result easily verified by direct integration of (7.4). By differentiation of $\phi_d(x, 0)$ we obtain the velocity $u_d(x, 0)$ due to the sources on the disc:

$$u_d(x, 0) = 2\pi x \int_0^b \frac{\eta q(\eta) d\eta}{(x^2 + \eta^2)^{3/2}}. \tag{10.6}$$

An alternative formula for u_d can be obtained in the following way. Assuming that the density $q(\rho)$ is differentiable, we have, by integration by parts,

$$\phi_d(x, 0) = -2\pi [q(b)\sqrt{x^2 + b^2} - q(0) | x |] + 2\pi \int_0^b q'(\eta) \sqrt{x^2 + \eta^2} d\eta. \tag{10.7}$$

Differentiating, we obtain

$$u_d(x, 0) = 2\pi \left[\mp q(0) - \frac{q(b)x}{\sqrt{x^2 + b^2}} + x \int_0^b \frac{q'(\eta) d\eta}{\sqrt{x^2 + \eta^2}} \right]. \tag{10.8}$$

The upper sign holds for $x < 0$, the lower for $x > 0$. Furthermore,

$$\frac{du_d}{dx} = 2\pi \left[-\frac{q(b)b^2}{(x^2 + b^2)^{3/2}} + \int_0^b \frac{q'(\eta) d\eta}{(x^2 + \eta^2)^{3/2}} \right]. \tag{10.9}$$

Assuming that the density q is a positive and non-increasing function of ρ, we see that $|u_d(x, 0)|$ is a decreasing function of $|x|$.

We now turn to the stream function. Replacing b in (8.1), and (8.2) and (8.6) by a variable of integration η, multiplying by $q(\eta)d\eta$ and integrating from 0 to b, we obtain

$$\psi_d(x, \rho) = \mp 2\pi\rho \int_0^\infty e^{-|x|s}\chi(s)J_1(\rho s)ds \begin{cases} +0 & \text{for } x \leqq 0, \\ -2m & \text{for } x \geqq 0, \end{cases} \tag{10.10}$$

where $\chi(s)$ denotes the function defined by (10.3), while m is the total strength of the disc, given by (10.1). The stream function (10.10) is single-valued and continuous in the domain D_b, defined in Sec. 8. The value of ψ_d on the streamline $x=0$, $\rho>b$ is obtained by multiplying the corresponding value $-2\pi\eta$ for a circumference of radius η by $q(\eta)d\eta$ and integrating the product from 0 to b. Since the integral in question is by (10.10) equal to the total strength of the disc, we have

$$\psi_d(0, \rho) = -m \quad \text{for} \quad \rho > b. \tag{10.11}$$

The value of the stream function for a disc on the streamline $x=0$, $\rho>b$, is equal to $-m$, independently of its radius b and of the density function $q(\rho)$.

This value being given by the right hand side of (10.10) for $x\to-0$, we have the identity

$$\rho\int_0^\infty \chi(s)J_1(\rho s)ds = \int_0^b \eta q(\eta)d\eta \quad \text{for} \quad \rho > b \geqq 0 \tag{10.12}$$

which could have been obtained directly by integrating the formula of Weber-Schafheitlin.

The results of this paragraph could be easily generalized by taking discontinuous density functions $q(\rho)$, possessing a finite or even an infinite number of jumps. As a particular case, we would obtain the potential and the stream function for one or several concentric rings of finite width.

11. Single-valued and many-valued Stokes' stream functions. As we have already mentioned, the domain D of definition of any Stokes' stream function is a subdomain of the upper half-plane $-\infty<x<+\infty$, $\rho\geqq 0$. It consists of all points in which the flow under consideration is regular. Sources, sinks and other singularities, as well as the x axis, are on the boundary of D. The Stokes' function has been defined by its total differential. It is therefore to be expected that this function will be many-valued when D is not simply connected. This is, for instance, the case for a circumference of sources, the corresponding domain D being bounded by the x-axis and by the point $x=0$, $\rho=b$, representing the trace of the circumference in the upper half-plane. On the other hand, the domain $D=D_b$ (see Sec. 8) corresponding to a disc of sources is simply connected. In fact, its boundary is a single continuum consisting of the x axis and of the vertical segment $0\leqq\rho\leqq b$. For this reason the stream function for a disc is necessarily single-valued, which is in agreement with the results of Sec. 10. For the same reason, any distribution of sources on the x axis will generate a single-valued stream function. The exclusive use of such sources has up to now prevented the recognition of an essential property of the Stokes' function.

In diagrams as usually given, the streamlines of an axially symmetric flow are drawn in the entire meridian plane, the lower half being the reflection of the upper one. In this representation, the domain corresponding to the simplest case of a point

source at the origin appears to be doubly connected, as is the case for a point source in plane flow. However, the corresponding stream function (3.2) is single-valued, while the stream function of the plane flow is many-valued, as is to be expected. Our explanation concerning the domain of definition, removes the paradox and our remarks apply to all cases when associated functions are defined by their total differential.

12. Blunt-nosed profiles. We turn now to the investigation of flows obtained by superposition of a parallel flow of constant velocity U in the direction of the positive x-axis and of a flow produced by an axially symmetric distribution of sources and sinks, around the axis.

As a first significant case, let us consider a single disc of (positive) sources in a parallel flow. The potential ϕ and the stream function ψ of the combined flow is given by the principle of superposition. Using (2.5), (10.2) and (10.10) we have

$$\begin{aligned} \phi &= Ux + \phi_d, \\ \psi &= \tfrac{1}{2}U\rho^2 + \psi_d. \end{aligned} \tag{12.1}$$

Let us now consider a *family of discs of variable* radius b with their centers at the origin and their axes coinciding with the x axis, all these discs having the *same total strength* m, independently of b. For $b=0$, when the total strength m is concentrated in a point-source at the origin, the dividing profile $\psi=0$ will be the classical Blasius-Fuhrmann half-body. As b increases from zero, we will have for sufficiently small values of this parameter a new family of profiles of half-bodies intersecting the ρ-axis above the edge of the corresponding disc.

In order to obtain this point of intersection which will give us the radius of the *main parallel*, $x=0$, of the half-body, we have to solve the equation

$$\psi(0, \rho) = 0. \tag{12.2}$$

Since its solution ρ is, by assumption, greater than b, we obtain, from (12.1) and (10.11),

$$\tfrac{1}{2}U\rho^2 - m = 0.$$

This equation yields the following result: *The radius $\rho=\rho_0$ of the main parallel is given by*

$$\rho_0 = \sqrt{\frac{2m}{U}}. \tag{12.3}$$

We see that ρ_0 *is independent of the radius b of the disc, which cannot exceed ρ_0.*

The velocity components u and v are obtained from (2.1) by (2.4). On the negative x axis we have $v=0$ and

$$u(x, 0) = U + u_d(x, 0), \tag{12.4}$$

the term $u_d=\partial\phi_d/\partial x$ being negative. A general theorem on the normal derivative of a surface distribution (Kellogg)[10] yields for $x=-0$ the result

$$u(-0, 0) = U - 2\pi q(0). \tag{12.5}$$

[10] O. Kellogg, *Potential theory*, J. Springer, Berlin, 1929, p. 164, theorem VI.

(More generally, we have $u(-0, \rho) = U - 2\pi q(\rho)$ for $0 \leqq \rho < b$. Observe that the integral in Kellogg's theorem VI vanishes identically for a disc.) Let us now assume that the (positive) density $q(\rho)$ is a non-increasing function of ρ. According to Sec. 10, the velocity u is then negative on the x axis and increases steadily in absolute value as x increases from $-\infty$ to 0. At the same time the total velocity $u = U + u_d$ decreases steadily from the value U to $U - 2\pi q(0)$ and remains therefore different from zero if $U - 2\pi q(0) > 0$. We have the following result: *the necessary and sufficient condition for the existence of a single stagnation point on the negative x axis is given by the inequality*

$$U \leqq 2\pi q(0). \tag{12.6}$$

Since $q(\rho)$ is non-increasing, the value $q(0)$ is not less than $m/\pi b^2$ which represents the constant density of a homogeneous disc with the same radius b and the same total strength m. Therefore, there will certainly be a single stagnation point for

$$U \leqq 2\pi m/\pi b^2 = 2m/b^2. \tag{12.7}$$

Considering U and m as given, we find that *a stagnation point* x_N *exists for* $b \leqq (2m/U)^{1/2}$. On the other hand, we have seen that a dividing line, $\psi = 0$, cannot exist for $b > (2m/U)^{1/2}$. *To every disc of radius* $b \leqq (2m/U)^{1/2}$ *corresponds a profile of a half-body*, the radius of the main parallel being $\rho_0 = (2m/U)^{1/2}$ independently of b.

Let $q_1(\rho)$ denote the density for a disc of radius one. Let us define the density for a disc of radius b by the formula

$$q(\rho) = \frac{1}{b^2} q_1\left(\frac{\rho}{b}\right), \qquad 0 \leqq \rho \leqq b. \tag{12.8}$$

The total strength is obviously independent of b. According to (12.4) and (10.6), the velocity on the x axis is given by

$$u = U + 2\pi x \int_0^b \frac{\eta q(\eta) d\eta}{(x^2 + \eta^2)^{3/2}} = U + 2\pi x \int_0^1 \frac{\xi q_1(\xi) d\xi}{(x^2 + b^2\xi^2)^{3/2}}. \tag{12.9}$$

This formula shows that for any fixed negative value of x, the velocity u is steadily increasing with b. We see that *the distance* $|x_N|$ *of the stagnation point from the origin is* steadily decreasing with increasing b. As $b \to \rho_0$ its abscissa x_N tends to a limit value, which will be negative for $U - 2\pi q(0) > 0$, and zero for $U - 2\pi q(0) = 0$. (q denotes here the density for the disc of radius ρ_0). For a non-increasing density, the second case occurs only for a disc of uniform density $m/\pi\rho_0^2$.

A distribution of sources over a finite region acts at infinity as the so-called equivalent point source at the origin with the same total strength m. For this reason, all our profiles are half-bodies with the same asymptotic radius

$$\rho_\infty = \sqrt{2}\, \rho_0 \tag{12.10}$$

which has been computed in (5.5), for the limiting case of the Blasius-Fuhrmann half-body.

We summarize now our results, which, in their entirety, hold for a non-increasing positive density given by the distribution law (12.8). We have obtained a *family of half-bodies depending on a parameter b. The only member of this family mentioned in the existing literature, is the Blasius-Fuhrmann half-body corresponding to the limiting*

case $b=0$. As b increases from 0 to $\rho_0=(2m/U)^{1/2}$, we obtain the other members of the family which are (increasingly) blunt-nosed and tend, for $b\to(2m/U)^{1/2}$, to a limiting profile. All these profiles have the same main parallel of radius ρ_0 and the same asymptotic radius at infinity. Similar results will be given later for other symmetric flows in three as well as in two dimensions.

In the following section, we shall turn to the investigation of some particular cases which we shall call the *integrable cases*. A computer would doubtless like to be given some examples, at least, in which the formula (10.10) for the stream function could be simplified by a convenient choice of the function $\chi(s)$. Unfortunately, $\chi(s)$ is connected with the density q by the integral equation of the first kind (10.3), and cannot be taken arbitrarily. So we have to take the density as the arbitrary function and discuss the *integrable cases*, in which the formula (10.3) can be simplified. Let us also point out that the distribution of sources is actually more important than the function $\chi(s)$, since it gives us at least a qualitative idea of the shape of the profile.

13. Discs of uniform density. In this case we have to take $q=m/\pi b^2$. By (10.3) we have

$$\chi(s) = \frac{m}{\pi b^2}\int_0^b \eta J_0(\eta s)\,d\eta.$$

Setting $s\eta=\xi$ and using the classical formula

$$\xi J_0(\xi) = \frac{d}{d\xi}\{\xi J_1(\xi)\}, \tag{13.1}$$

we obtain

$$\chi(s) = \frac{m}{\pi b s} J_1(bs). \tag{13.2}$$

By (12.1), (10.2) and (10.10), the explicit formulae for ϕ and ψ are therefore given by

$$\phi(x,\rho) = Ux - \frac{2m}{b}\int_0^\infty e^{-|x|s} J_0(\rho s) J_1(bs)\frac{ds}{s}, \tag{13.3}$$

$$\psi(x,\rho) = \tfrac{1}{2}U\rho^2 \mp \frac{2m\rho}{b}\int_0^\infty e^{-|x|s} J_1(\rho s) J_1(bs)\frac{ds}{s}\begin{cases} +0 & \text{for } x \leqq 0 \\ -2m & \text{for } x \geqq 0. \end{cases} \tag{13.4}$$

The profile is given by the equation $\psi=0$. The terms containing the integrals have been denoted in Sec. 10 by ϕ_d and ψ_d, respectively. It is obvious that these functions could have been obtained directly by integration of the potential and stream functions for a circumference. Using the general equation (10.11) ,we obtain the formula

$$\int_0^\infty J_1(\rho s) J_1(bs)\frac{ds}{s} = \frac{1}{2}\frac{b}{\rho} \quad \text{for} \quad \rho > b \tag{13.5}$$

which is a special case of the Weber-Schafheitlin discontinuous integral [Watson,[7] p. 405 (1)] but which has been proved here directly from the basic principles of Potential Theory.

The velocity components are given by

$$u(x, \rho) = U \mp \frac{2m}{b} \int_0^\infty e^{-|x|s} J_0(\rho s) J_1(bs) ds, \tag{13.6}$$

$$v(x, \rho) = \frac{2m}{b} \int_0^\infty e^{-|x|s} J_1(\rho s) J_1(bs) ds. \tag{13.7}$$

The upper sign holds for $x \leqq 0$, the lower, for $x \geqq 0$. The stagnation point is obtained by putting $u(x, 0) = 0$ in the first Eq. (13.6). According to (12.9), this gives

$$U + \frac{2mx}{b^2} \int_0^b \frac{\eta d\eta}{(\eta^2 + x^2)^{3/2}} = 0; \qquad (x \leqq 0).$$

The integral is this formula can be easily computed. In this way we obtain, for $x = x_N$,

$$U - \frac{2m}{b^2}\left(1 + \frac{x_N}{\sqrt{x_N^2 + b^2}}\right) = 0.$$

Observing, that $2m/U = \rho_0^2$, by (12.3), we see that

$$x_N^2 = \frac{b^2(\rho_0^2 - b^2)}{\rho_0^4 - (\rho_0^2 - b^2)^2}. \tag{13.8}$$

According to the general results of Sec. 12, the distance $|x_N|$ of the stagnation point from the disc takes, for $b = 0$, the value $|x_N| = \rho_0/\sqrt{2}$, corresponding to the Blasius-Fuhrmann half-body [*cf.* (15.5)], and decreases steadily to zero, as $b \to \rho_0$. This fact can be verified in an elementary way by using (13.8). The radius of the main parallel is $\rho_0 = \sqrt{2m/U}$; the asymptotic radius is $\rho_\infty = \rho_0\sqrt{2}$. The nose of the profile becomes blunter with increasing b. The value $b = \rho_0$ corresponds to a limiting singular case of an *ultra-flat profile** with a nose coinciding with the disc. The velocity in this case becomes infinite at the edge of the disc.

14. Curvature of the profiles. We have seen in Sec. 5 that the curvature K of the Blasius-Fuhrmann half-body (corresponding to $b = 0$) is steadily decreasing along the profile. For the other limiting case ($b = \rho_0$) the curvature of the nose of the ultra-flat profile is identically zero along the disc, but jumps suddenly to infinity at its edge. Since changes take place gradually, we may venture the following *conjectures about the curvature K of the intermediate profiles.*

As b increases from zero, the curvature of the profile will decrease in the vicinity of its stagnation point and will increase in the vicinity of the edge b. There will be a moment in which we will have an *equalization of curvatures* on a certain arc adjacent to the stagnation point. At this moment the profile will have a *nearly-spherical cap*, while the adjacent infinite branch of the dividing line will have a curvature steadily decreasing to zero. This stage can last for awhile, but, as b continues to increase, the edge of the disc, loaded with sources, will act like a spearhead repulsing the parts of

* Professor G. Birkhoff has kindly drawn my attention to the fact that the *potential* of the flow around the ultra-flat profile has been computed (in the neighborhood of the nose) by T. L. Smith in 1943.

the profile in its immediate neighborhood. There will appear a maximum of K in the vicinity of this edge, while the nose continues to become blunter. As b increases still further, this maximum will become greater and its ordinate will tend to ρ_0. The final stage will be reached when b takes the value ρ_0 at which moment this maximum becomes infinite.

15. Discs with Bessel's distribution of sources. Let us first consider a disc of unit radius. In order to obtain another integrable case, we take as surface density the function

$$q_1(\rho) = C_1 J_0(j\rho), \tag{15.1}$$

where $j = 2.40483$ denotes the smallest positive root of $J_0(z)$, so that $q(1) = 0$. (This is the most interesting case. However, the following consideration would hold with slight modifications for other positive values of the constant j.) The total strength m is connected with the constant C_1 by the equation

$$2\pi C_1 \int_0^1 \eta J_0(j\eta) d\eta = m.$$

Using the classical formula (13.1), we obtain

$$C_1 = \frac{mj}{2\pi J_1(j)}. \tag{15.2}$$

It is well known that $J_1(j)$ is different from zero. For $m > 0$, the density $q_1(\rho)$ is positive and decreasing in the interval $0 \leqq \rho \leqq 1$.

Let us now consider a disc of radius b and of the same strength m. According to (12.8), we put

$$q(\rho) = \frac{mj}{2\pi b^2 J_1(j)} J_0\left(\frac{j\rho}{b}\right) \tag{15.3}$$

and call this formula *Bessel's distribution law.*

Let us introduce the abbreviations

$$C = \frac{mj}{2\pi b^2 J_1(j)}, \qquad \alpha = \frac{j}{b}. \tag{15.4}$$

By (15.3), we have then

$$q(\rho) = C J_0(\alpha\rho). \tag{15.5}$$

The integral in (10.3) can be computed by using the following fundamental formula of the theory of Bessel functions

$$(\alpha^2 - s^2) \int_0^b \eta J_0(\alpha\eta) J_0(s\eta) d\eta = \left[\eta \left\{ s J_0(\alpha\eta) J_0'(s\eta) - \alpha J_0(s\eta) J_0'(\alpha\eta) \right\}\right]_{\eta=1}^{\eta=b}, \tag{15.6}$$

which gives for (10.3)

$$\chi(s) = \frac{C\alpha b}{\alpha^2 - s^2} J_0(bs) J_1(\alpha b) = \frac{mj^2}{2\pi(j^2 - b^2 s^2)} J_0(bs). \tag{15.7}$$

We remark that the function $\chi(s)$ remains finite and continuous for $s=j/b$, since $J_0(j)=0$.

Inserting the expression (15.7) in the general formulae of Sec. 10 for a disc and adding a uniform flow, we obtain the formulae for the resulting flow. From Sec. 12 we know that there is exactly one stagnation point x_N on the negative x axis for every value of $b<\rho_0$ and that x_N steadily tends to a negative limiting value with $b\to\rho_0$. All profiles pass through the point $x=0$, $\rho=\rho_0$ and all possess the same asymptotic radius $\rho_\infty=\rho_0\sqrt{2}$.

Let us consider the limiting profile (corresponding to $b=\rho_0$) which passes through the edge of the disc at the point $\rho=\rho_0$ of the ρ-axis. The density of the sources being zero at that point, we can extend the definition of $q(\rho)$ to a greater disc by putting $q(\rho)=0$ for $\rho>\rho_0$. The point $\rho=\rho_0$ is now an interior point of the greater disc in which the density is continuous and satisfies a Hoelder condition. It follows from well known theorems on tangential and normal derivatives of a surface distribution (Kellogg,[10] p. 162, theorem V and p. 164, theorem VI) that the velocity components u and v remain continuous for $x=0$, $\rho=\rho_0$, and that the limiting profile has, at the point in question, a continuous tangent. However, the curvature of the profile becomes infinite at the same point. (All these results can be easily checked by the formulae given in this paragraph, provided that certain precautions are taken in the use of discontinuous integrals.) Our limiting profile is non-analytic at the point ρ_0. We note that a *singularity on the profile* can be obtained only when the profile passes through the boundary of the region occupied by the sources.

SOME GENERAL RESULTS IN THE THEORY OF NOISE THROUGH NON-LINEAR DEVICES*

BY
DAVID MIDDLETON
Harvard University

1. Introduction. Because of the great prevalence of noise in almost all electronic processes, a study of the nature and properties of such noise seems desirable. The term "properties of noise" is used here to indicate such measurable quantities as the average and the mean-square voltages and currents, the mean power spectrum, the power or energy associated with the wave, and the power and the correlation function of the disturbance, or of part of it, when noise or a signal and noise is modified by passage through non-linear apparatus. From an analytical point of view, the theory of noise is intrinsically related to that of the Brownian motion so that the results in the discussion of the one may bear rather closely upon the other. There are two different but equivalent lines of attack on the over-all problem: in noise theory we are primarily interested in what happens to random noise waves (with or without an accompanying signal) when they are passed through non linear devices, such as second detectors or mixers in radio receivers, for example, or amplifiers in which cutoff and/or overloading is present. Here the *Fourier series method of Rice*,[1,2] is the more natural approach and is the one followed in the present paper. In the study of Brownian motion and fluctuation phenomena in general, where the variations in the system are described by a diffusion process, the second method of *Fokker-Planck*, or *the diffusion equation method*, is used. We shall not consider this approach here; an excellent discussion has recently been given by Wang and Uhlenbeck,[3] and less recently, an interesting treatment of somewhat similar subjects by Chandrasekhar.[4] We mention only in passing that the two methods can be shown to yield identical results.[3]

Rice† and others[5–9] have used the Fourier series method in the solution of special

* Received Jan. 29, 1947.

[1] S. O. Rice, *Mathematical analysis of random noise*, Bell Sys. T. J. **23**, 282 (1944).

[2] S. O. Rice, loc. cit. **24**, 46 (1945). Since this paper was written, Professor Brillouin has kindly called the author's attention to the interesting work of Dr. Blanc-Lapierre along somewhat similar lines, in particular, a thesis: *Sur certaines fonctions aléatoires stationnaires. Application à l'étude des fluctuations dues à la structure électronique de l'électricité*, and *Effet Schottky. Fluctuations dans les amplificateurs linéaires et dans les détecteurs*, Bull. de la Soc. fr. des Elec. (6) **5**, No. 53, Nov. 1945.

[3] M. C. Wang and G. E. Uhlenbeck, *On the theory of the brownian motion II*, Rev. Mod. Phys. **17**, 323 (1945). This paper and reference 4 contain a considerable number of references to previous work along these lines.

[4] S. Chandrasekhar, *Stochastic problems in physics and astronomy*, Rev. Mod. Phys. **15**, No. 1, 1, 1943.

† See part IV of ref. 2.

[5] K. Fränz, Zeits. f. Hoch. u. Elek. **57**, 146 (1941).

[6] W. R. Bennett, *The response of a linear rectifier to signal and noise*, J. Amer. Acous. Soc. **15**, 165 (1944).

[7] D. O. North, *The modification of noise by certain non linear devices*, a synopsis of which was presented at the Jan. 1944 winter technical meeting of the I.R.E.

[8] J. H. Van Vleck and D. Middleton, *A theoretical comparison of the visual, aural, and meter reception of pulsed signals in the presence of noise*, J. Appl. Phys. **17**, 940 (1946).

[9] D. Middleton, *The response of biased, saturated linear and quadratic rectifiers to random noise*, Jour. A. Phys. **17**, 778 (1946).

problems involving the rectification of noise, or of a signal and noise. It is the purpose of this paper to generalize some of the results of previous work on this topic and to obtain original results for a number of unsolved problems in the analysis of noise through non-linear devices. Specifically, it is believed that the treatment of the following topics is new and of interest.

(a) Passage of a *modulated* signal in the presence of noise through a general non-linear apparatus. The case of a sinusoidally modulated carrier (Sec. 2), is examined and attention is also given to the case of narrow-band noise, symmetrically distributed in frequency about the carrier (Sec. 3).

(b) The biased νth-law rectifier, for modulated and unmodulated carriers (Sec. 4). Limiting cases of large noise or signal voltages are also discussed. Reference 9 gives a detailed discussion of this problem for linear and quadratic rectifiers where noise alone is rectified.

(c) The problem of a modulated signal and narrow-band noise, with a determination of the various probability densities associated with the envelope of the wave (Sec. 5). Section (c) and (d) offer alternative solutions to some of the problems discussed in (a) and (b).

(d) The correlation function and mean power associated with the envelope of signal and noise. Attention is given to the low frequency output of the half-wave νth-law device (Sec. 6).

(e) The νth-law, half-wave rectification of noise alone, the results of which are of interest in the measurement of noise by meters, spectrum analyzers, etc., and in the detection of pulse signals in the presence of noise[8,9] (Sec. 7). This work is a generalization of reference 9 in that ν can take on any positive value, but is less general in that only half-wave detection is treated.

(f) A general "small-signal" theory, in which the speak values of the incoming wave, whether noise or a signal and noise, are sufficiently small that overloading and cut-off do not occur. Rectification takes place because of the curvature of the dynamic characteristic of the device in question.

Not all the material in the present paper is original, it is realized, but in the discussion of (a)–(f) it has been necessary for clarity of treatment to bring together and extend, when necessary, a number of results previously derived in probability theory which are fundamental and hence unavoidable in the analysis of problems of this type. These results include the generalized, s-dimensional random-walk problem,[4] from which in turn one may obtain the characteristic function, with the distinctive property of being the Fourier transform of the probability density, and finally the central-limit theorem,[10] which in the limit of a very large number of events can be shown to yield the s-dimensional Gaussian distribution characteristic of all random processes fulfilling certain rather elastic conditions with regard to the separate distributions of the various events. Some of the details are available in Appendix I. Furthermore, although the concept of the correlation function and its relation to the mean power spectrum has been examined in varying detail elsewhere, we include a brief treatment in Appendix II, along with some of the more significant properties of the correlation function which are necessary in our work. Rigor has not been preserved at all costs, *vide* the use of the Dirac delta-function; the physical significance

[10] H. Cramèr, *Random variables and probability distributions*, Cambridge Tract No. 36 (1937), Chapters VI and X. References 1 and 2 also contain references to this problem; see Secs. 2.9 and 2.10.

of the results is used to check any possible weakness in the rigor. Finally, in Appendix III some of the more unfamiliar special functions that appear in our results are briefly considered. It should be kept in mind that the present discussion applies only when the dynamic path of the non linear device is a one-valued function of the input disturbance. When the path is multi-valued, the theory breaks down. For example, when the plate-load of a rectifying tube is a pure resistance, or at worst, is primarily resistive, the dynamic path is one-valued, or nearly so. Figure 1 illustrates a typical tube characteristic under these conditions. However, when there is appreciable reactance X_p in comparison with the resistance R_p, the problem is not tractable with the present methods.

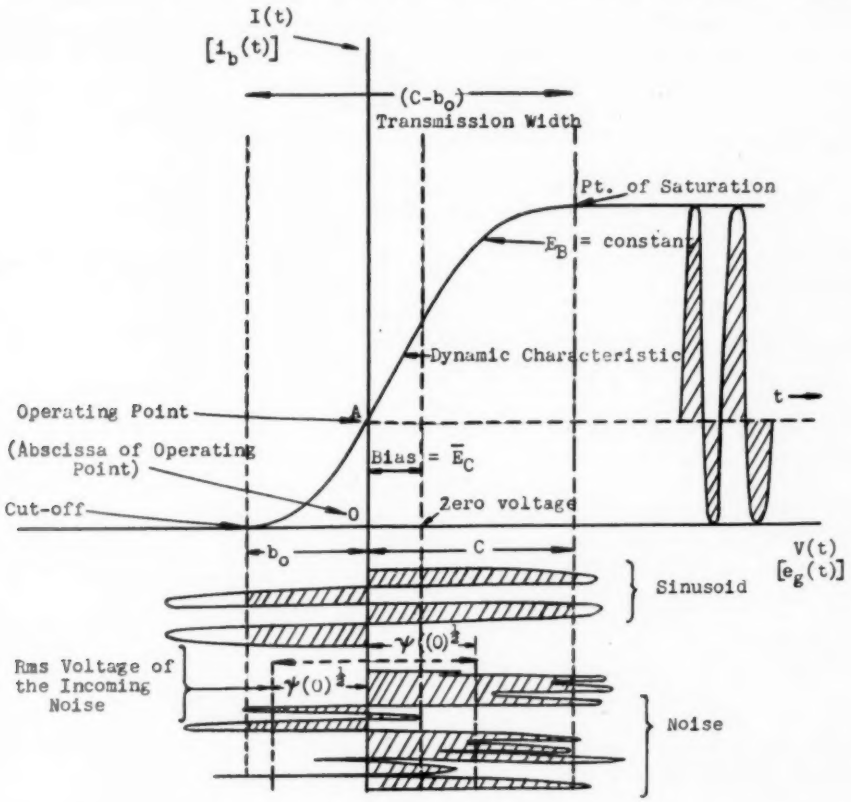


FIG. 1. Typical dynamic characteristic for resistive loads.

At this point it is well to mention what we mean by a random process, and with what class of such processes problems of the present kind deal. We consider $X(t)$ to be a random process when $X(t)$ does not depend in a definite way on the independent variable t, but instead may be specified by an aggregate of different functions $X(t)$, of which experimentally only probability distributions are observable. Here $X(t)$ may be the displacement of the electron beam on the screen of a cathode ray oscilloscope,

or, say, the velocity of a particle in Brownian motion. The quantity $X(t)$ may also represent a combination of two or more such stochastic variables, in which case we may speak of a two- or more-dimensional random process. In what follows, we shall consider formally that X and t are continuous variables; for our work, based as it is on the problem of the random walk (see Appendix I), this a plausible assumption. Now, as Wang and Uhlenbeck[3] have shown, it is possible to describe a random process completely by means of a set of distributions determining the probability that X lie in the range X, $X+dX$, at time, t, that X_1 and X_2 fall in the intervals X_1, X_1+dX_1 and X_2, X_2+dX_2, at times t_1 and t_2, respectively, and so on. The classification is considerably simplified when the initial times of the observations do not enter; processes of this kind are *stationary*. Their statistical properties are independent of when the measurements are made, and depend only on the duration of the observations. There is an important class in this group known as *Markoff processes*, which are completely described by the second order density $W_2(X_1, X_2; t)$:

$$W_2(X_1, X_2; t) = \text{joint probability of obtaining } X_1 \text{ in the range } X_1, X_1+dX_1 \text{ and } X_2 \text{ in the range } X_2, X_2+dX_2 \text{ at a time } t \text{ later.} \tag{1.1}$$

Since the present paper deals exclusively with distributions that are stationary and Markoffian, or if not originally the latter, can be extended to a more complex Markoff process by the introduction of suitable additional random variables,[11] we find it helpful to mention the properties of W_2. We have

$$W_2(X_1, X_2; t) = W_1(X_1)P_2(X_1 \mid X_2; t), \tag{1.2}$$

where following Wang and Uhlenbeck,[3] $P_2(X \mid X_2; t)$ denotes the *conditional* probability that, given X_1, we find X_2 in the interval X_2, X_2+dX_2 a time t later. Here $W_1(X_1)$ is simply the first order distribution, giving the probability of locating X_1 in the interval X_1, X_1+dX_1; the time does not enter because the process is assumed to be stationary. We have also the relations

$$\begin{gathered} P_2(X_1 \mid X_2; t) \geqq 0; \qquad \int dX_2 P_2(X_1 \mid X_2; t) = 1; \\ W_1(X_2) = \int W_1(X_1)P_2(X_1 \mid X_2; t)dX_1, \end{gathered} \tag{1.3}$$

where the region of integration includes all possible values of the variables. The further condition

$$\begin{aligned} P_n(X_1, t_1; X_2, t_2; \cdots; X_{n-1}, t_{n-1} \mid X_n; t_n) \\ = P_2(X_{n-1}, t_{n-1} \mid X_n; t_n), \qquad t_n > t_{n-1} > \cdots > t_1 \end{aligned} \tag{1.4}$$

shows that (1.1), or P_2, completely determines the Markoff process, for once W_2 is known, we may obtain all distributions W_n, $n \geqq 2$, from the above. It should be observed that $P_2 \to W_1(X_2)$ as $t \to \infty$, provided X_2 is purely random and does not contain periodic components. As has been previously pointed out,[3] W_2 or P_2 cannot be selected arbitrarily, for it must satisfy the fundamental equation

[11] See, for example, sec. 3 (c), of ref. 3.

$$P_2(X_1 \mid X_2; t) = \int dX P_2(X_1 \mid X; t_0) P_2(X \mid X_2; t - t_0), \qquad 0 \leqq t_0 \leqq t. \tag{1.5}$$

which is due to Smoluchowski. Equations (1.1)–(1.5) have been written for one- and two-dimensional systems; the extension to a greater number follows at once without modification of the concepts.

The successful solution of non-linear problems involving noise by the Fourier series method of Rice depends, then, on the following definitions, devices, and assumptions.

(1) It is *assumed* that the random process is stationary and Gaussian, i.e., X_1, X_2, etc., all obey a Gaussian distribution law. In the completely Gaussian case, all components of X_1, X_2, $\cdots$, as well, are postulated to have a normal distribution of amplitudes. For our purposes, however, this is not necessary, for by the central limit theorem X_1, X_2, etc., will be Gaussian, irrespective of the distribution law of their components as long as the latter are sufficiently numerous, under conditions easily fulfilled physically in most instances. For convenience, we shall assume that our random processes are initially entirely random in the normal sense. In either case $W_2(X_1, X_2; t)$ is given by (A1.3).

(2) Our choice of input spectra will determine whether or not the system is Markoffian.[12,13] If it is, then W_2 completely describes it; however in the present paper it is not usually critical that we be able to classify the process thoroughly; the chief concern is with the amplitude distribution and with the mean power spectrum, which may always be found when W_1 and W_2 are known, and here it is sufficient that the system be stationary.

(3) The mean power spectrum $W(f)$ may be obtained directly from the correlation function $R(t)$, defined by Eq. (A2.3), since by a well-known result[14–16] they are each other's cosine Fourier transforms:

$$W(f) = 4\int_0^\infty R(t) \cos \omega t dt; \quad \omega = 2\pi f; \quad R(t) = \int_0^\infty W(f) \cos \omega t df. \tag{1.6}$$

The correlation function, in turn, follows from the definition (A2.3) and with the help of W_2 is seen to be

$$R(t) = \iint_{-\infty}^{\infty} X_1 X_2 W_2(X_1, X_2; t) dX_1 dX_2. \tag{1.7a}$$

If the correlation function for the random wave $X(t)$ *after* passage through a non linear device is desired (1.7a) becomes

$$R(t) = \iint_{-\infty}^{\infty} g(X_1) g(X_2) W_2(X_1, X_2; t) dX_1 dX_2, \tag{1.7b}$$

---

[12] J. L. Dobb, Ann. Amer. Stat. **15**, 229 (1944).

[13] Reference 3, Sec. 7, and note II of the appendix for details.

[14] N. Wiener, Acta Math. **55**, 117 (1930).

[15] A. Khintchine, Math. Ann. **109**, 604 (1934).

[16] G. I. Taylor, Proc. Lond. Math. Soc. Sec. 2, **20**, 196 (1920), and Proc. Royal Soc. **164**, 476 (1938) gives applications to the theory of turbulence. Rice's interesting papers [1,2] treat a large number of noise problems with the help of (1–6), while ref. 3 puts more emphasis on the Brownian motion.

where $g(X)$ is the dynamic characteristic of the apparatus in question, and X_1 is the incoming disturbance at some initial time and X_2 at time t later.

(4) The explicit evaluation of the integral (1.7b) is often very difficult in this form because discontinuities in the characteristic g appear as finite limits in the integration. Then it is convenient, almost mandatory, to introduce the *Fourier transform* $f(iz)$ of the dynamic path, so that the output wave is given in terms of the input by[17]

$$g(X) = \frac{1}{2\pi}\int_{\mathbf{C}} f(iz)e^{iXz}dz. \tag{1.8}$$

The contour **C** extends from $-\infty$ to $+\infty$ along the real axis and is indented downward about a pole or branch point at the origin. The evaluation of (1.8) is effected when the contour is extended in an infinitely large semi-circle in a counterclockwise (positive) or clockwise (negative) sense, depending on whether the coefficient of iz in the exponent is positive or negative, respectively. When **C** is traversed in a positive direction, the residue at $z=0$ yields the output as the desired function of the input, while for a negative circuit of the contour, $g(X)$ vanishes. Other contours are also possible, and combinations of such paths may be used for complicated characteristics. In this way, we are able to distinguish between the transmission and the cutoff states of the apparatus. It should be mentioned that when the contour representation (1.8) is employed, the results for the correlation function (1.7b) will be expressed in terms of *the characteristic function* associated with W_2, namely Eq. (A1.1), in the general s-dimensional case, rather than by W_2 itself. This is not surprising, for it is easily shown that the characteristic function is the Fourier transform of the corresponding probability density. See Eqs. (2.13) to (2.16), in the next section.

(5) When a signal, modulated or unmodulated, is introduced along with the noise, the above concepts and artifices hold, with slight modification. For instance, (1.7a, 7b) now represent only the correlation function due to the noise components in the wave; a further average or averages are necessary to account for the signal and modulation components. Here the average over the phases of these components must be performed. In the next section this extension of the method is outlined and illustrated (see also references 2, 6, and 8).

2. Rectification of modulated signals in the presence of noise. Before determining what happens when a modulated wave is detected in the presence of noise, we must represent the incoming disturbance analytically. Let us consider first the signal component, which we shall denote by $V_s(t, t')$. Here the variable t refers to the time-variations of the carrier and t' to those of the modulation. Physically, t and t' both represent the same instant in time, our choice of symbols being merely one of mathematical convenience. For the type of signal of greatest interest here we may therefore write

$$V_s(t, t') = A_0(t')\cos\omega_0 t, \qquad \omega_0 = 2\pi f_0, \qquad A_0(t') \geqq 0, \tag{2.1}$$

where f_0 is the carrier frequency and $A_0(t')$ is some function of the time, which is only properly called the modulation when it varies slowly compared with $\cos\omega_0 t$. In general the carrier and modulation are not commensurable and hence are uncorre-

[17] W. R. Bennett and S. O. Rice, Phil. Mag. **18,** 422 (1934). For an extensive application of the contour representation, see reference 9.

lated; t and t' are consequently independent variables. On the other hand, there are cases when $A_0(t')$ and $\cos \omega_0 t$ bear a commensurable relation to each other, correlation exists, and t and t' are then functionally related. Such instances may occur, for example, when the carrier is over-modulated.

Now let us consider briefly the noise portion of the disturbance entering the non linear device. Several satisfactory representations are possible,[18] of which we choose the Fourier series form

$$V_N(t) = \sum_{n=1}^{N} \{a_n \cos \omega_n t + b_n \sin \omega_n t\}, \qquad \omega_n = 2\pi f_n = 2\pi n/T, \tag{2.2}$$

where the interval of expansion lies between 0 and T, and N is very much greater than unity. The quantities a_n and b_n are independent random variables, having the following properties:

$$\overline{a_n} = \overline{b_n} = 0; \quad \overline{a_n b_m} = 0; \quad \overline{a_n a_m} = \overline{b_n b_m} = w(f_n)\Delta f \delta_n^m = \{|S(f)|^2/T\}\delta_n^m, \tag{2.3}$$

this last from (A2.4). The bar indicates the statistical average over the various random quantities (see Appendix II); here δ_n^m is the familiar Kronecker delta, which has the values $\delta_n^m = 0$, $m \neq n$; $\delta_n^n = 1$. We assume the random process describing the noise to be stationary; this assumption can be verified experimentally. The time average for $T \to \infty$ and the ensemble average, in which the average of an (indefinitely) large number of finite intervals or "strips" corresponding to separate observations is taken, then yield the same results. Under these circumstances the distribution of the random variables a_n and b_n is unaltered. It is convenient to assume also that this distribution is Gaussian, with the standard deviation $[w(f_n)\Delta f]^{1/2}$, but as the central limit theorem shows, [see Appendix I (A1.1, A1.2, A1.3)], any other distribution with the same average and second moments [Eq. (2.3)] leads to identical expressions for the noise wave. Equations (2.3) may then be interpreted as follows: the first relation shows that $\overline{V_N(t)} = 0$, the second shows that a_n and b_n are independent, and the third apportions the mean power in the frequency range $f_n, f_n + \Delta f_n$ dissipated by a current flowing in a unit resistance when a potential difference $V_N(t)$ is maintained between the terminals. In the limit of a very long time, or what is the same thing, of a very large number of "strips" in the ensemble, $\Delta f (= 1/T)$ becomes infinitesimal, and the summation may be replaced by an integral from $f = 0$ to $f = \infty$.

With the aid of Appendix I, we proceed to determine $W_1(X)$ and $W_2(X_1, X_2; t)$, cf. (1.1), where we identify V_N at some time t_0 with X_1 and some time t later with X_2. Referring to Eq. (A1.2), we find from (2.2) and (2.3) that on letting $X_{(j=n)} = a_n \cos \omega_n t + b_n \sin \omega_n t$,

$$\begin{aligned} \overline{X} &= \nu = \sum_{n=1} (\bar{a}_n \cos \omega_n t + \bar{b}_n \sin \omega_n t) = 0; \\ \mu_{11} &= \overline{X^2} = \sum_{n=1}^{N} (\overline{a_n^2} \cos^2 \omega_n t + \overline{b_n^2} \sin^2 \omega_n t) = \sum_{n=1}^{N} (\overline{a_n^2} + \overline{b_n^2})/2 \to \int_0^\infty w(f)df \equiv \psi(0), \end{aligned} \tag{2.4}$$

with $\psi(0) = \overline{X^2}$ the mean power in the wave. Clearly here the matrix $\mathfrak{y}$ associated with W_2, (see A1.3), has the single element $\psi(0)$ and $|\mu| = \psi^{1/2}$, $\mu^{11} = 1$, so that the first order distribution for $V_N = X$ is, from (A1.3),

[18] See reference 1, Sec. 2.8, and reference 6.

$$W_1(X) = [2\pi\psi(0)]^{-1/2} \exp[-X^2/2\psi(0)]. \tag{2.5}$$

In a similar fashion the joint probability density $W_2(X_1, X_2; t)$ follows; we have

$$\begin{aligned} &\overline{X}_1 = \overline{X}_2 = \nu_1 = \nu_2 = 0, \quad \text{and} \\ &\mu_{11} = \sum_{n=1}^{N} (\overline{a_n^2} \cos^2 \omega_n t_0 + \overline{b_n^2} \sin^2 \omega_n t_0) = \overline{X_1^2} = \psi(0) \\ &\mu_{22} = \sum_{n=1}^{N} (\overline{a_n^2} \cos^2 \omega_n (t_0 + t) + \overline{b_n^2} \sin^2 \omega_n (t_0 + t)) = \overline{X_1^2} = \psi(0), \end{aligned} \tag{2.6}$$

from (A1.2) and (A1.4), as a consequence of the conditions (2.3). For the off-diagonal moments one may write

$$\begin{aligned} \mu_{12} = \mu_{21} = \overline{X_1 X_2} &= \sum_{n=1}^{N} [\overline{a_n^2} \cos \omega_n t_0 \cos \omega_n (t_0 + t) + \overline{b_n^2} \sin \omega_n t_0 \sin \omega_n (t_0 + t)] \\ &= \sum_{n=1}^{N} (\overline{a_n^2} \cos^2 \omega_n t_0 + \overline{b_n^2} \sin^2 \omega_n t_0) \cos \omega_n t \\ &\to \int_0^\infty w(f) \cos \omega t df \equiv \psi(t). \end{aligned} \tag{2.7}$$

The quantity $\psi(t)$ is the correlation function of the input noise, and by (1.6),[19] the Fourier transform of the input spectrum $w(f)$. For the two-dimensional distribution μ becomes

$$\begin{aligned} &= \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} = \begin{bmatrix} \psi(0) & \psi(t) \\ \psi(t) & \psi(0) \end{bmatrix}; \quad |\mu| = \psi^2(1 - r^2(t)), \quad \text{and} \\ &\mu^{11} = \mu^{22} = \psi(0), \qquad \mu^{12} = \mu^{21} = -\psi(t); \end{aligned} \tag{2.8}$$

$r(t)$ is the normalized correlation function of the incoming disturbance, Eq. (A2.7). With the aid of (A1.3) the joint distribution is easily observed to take the familiar form[20]

$$W_2(X_1, X_2; t) = \frac{1}{2\pi\psi(1 - r^2)^{1/2}} e^{-(X_1^2 + X_2^2 - 2rX_1X_2)/2\psi(1-r^2)}. \tag{2.9}$$

The incoming wave is simply the sum of (2.1) and (2.2), with the various properties of the two components outlined above:

$$V_{in} = X + V_s(t', t). \tag{2.10}$$

Accordingly, if $g(V_{in})$ is the output of the device [see (1.7b)] the correlation function for the noise components is

$$R_N(t_0, t_0'; t) = \int\int_{-\infty}^{\infty} g(X_1 + V_s[t_0, t_0']) g(X_2 + V_s[t_0 + t, t_0' + t]) W_2(X_1, X_2; t) dX_1 dX_2. \tag{2.11}$$

---

[19] See Appendix II for details.

[20] Henceforth we abbreviate $\psi(0)$ by ψ.

The complete correlation function, which includes the average over the phases of the signal as well as over the noise, is

$$R(t) = T_0'^{-1}\int_0^{T_0'} dt_0' \int_0^{T_0} T_0^{-1} R_N(t_0, t_0'; t)dt_0, \tag{2.12}$$

where T_0 and T_0' are the respective periods of the carrier and the modulation. The double integration in (2.12) reduces to a single operation in case t_0 and t_0' are functionally related, i.e., when carrier and modulation have a constant phase difference. Since W_2 is given explicitly [Eq. (2.9)] from our assumptions regarding the character of the noise, (2.11) and (2.12) represent the formal solution to our problem, in as much as the mean power spectrum follows immediately from (1.6). However, as these equations in their present form can be handled only for a few special cases, chiefly in the instance of "small-signal" rectification, where the distortion of the dynamic path arising from cut-off and saturation is gradual, we make use of the device of contour integration, (1.8), for the large and important class of problems in which cut-off and saturation take place abruptly. Then the outputs at times t_0 and t_0+t are respectively

$$g(V_1) = g(V_s(t_0, t_0') + X_1) = \frac{1}{2\pi}\int_C f(iz)e^{iz(X_1+V_s[t_0', t_0'])}dz, \tag{2.13a}$$

$$g(V_2) = g(V_s[t_0 + t, t_0' + t] + X_2) = \frac{1}{2\pi}\int_{C'} f(i\xi)e^{i\xi(X_2+V_s[t_0+t, t_0+t])}d\xi, \tag{2.13b}$$

where $\mathbf{C}'$ is a contour similar to $\mathbf{C}$ but in the ξ-plane. Equations (2.11) and (2.12) become finally

$$\begin{aligned} R(t) &= \frac{1}{4\pi^2}\int_C f(iz)dz \int_{C'} f(i\xi)d\xi \left\{\int\int_{-\infty}^{\infty} W_2(X_1, X_2; t)e^{izX_1+i\xi X_2}dX_1dX_2\right\} \\ &\times \left\{\frac{1}{T_0T_0'}\int_0^{T_0}\int_0^{T_0'} dt_0'\,dt_0 e^{izV_s(t_0', t_0)+i\xi V_s[t_0+t, t_0'+t]}\right\} \\ &= \frac{1}{4\pi^2}\int_C f(iz)dz\int_{C'} f(i\xi)F_N(z, \xi, t)F_S(z, \xi; t)d\xi, \end{aligned} \tag{2.14}$$

in which $F_S(z, \xi; t)$ is by definition the characteristic function of the signal, given here by

$$F_S(z, \xi; t) = \frac{1}{T_0T_0'}\int_0^{T_0'} dt_0' \int_0^{T_0} dt_0 e^{izV_s(t_0, t_0')+i\xi V_s(t_0+t, t_0'+t)}, \tag{2.15}$$

and $F_N(z, \xi; t)$ is similarly seen to be the characteristic function for the noise, and may be written

$$\begin{aligned} F_N(z, \xi; t) &= \int\int_{-\infty}^{\infty} W_2(X_1, X_2; t)e^{izX_1+i\xi X_2}dX_1dX_2 \\ &= \exp\left[-\tfrac{1}{2}\psi(z^2 + \xi^2) - \psi(t)z\xi\right]. \end{aligned} \tag{2.16}$$

This last result is obtained from (A1.1) with the help of (2.6)–(2.9). From (2.14), it

appears that the characteristic function for a modulated signal and noise is merely the product of the separate characteristic functions, for physically the signal and noise components are uncorrelated, and hence their distributions must be independent.

For the modulated carrier Eq. (2.1) may be substituted directly in (2.15). The more usual case of no carrier-modulation correlation gives us

$$F_s(z,\xi;t) = \frac{1}{T_0'}\int_0^{T_0'} dt_0' \left\{\frac{1}{T_0}\int_0^{T_0} e^{izA_0(t_0')\cos\omega_0 t_0 + i\xi A_0(t_0'+t)\cos\omega_0(t_0+t)} dt_0\right\}$$

$$= \frac{1}{T_0'}\int_0^{T_0'} dt_0' \left\{\sum_{m=0}^{\infty}\sum_{p=0}^{\infty} \epsilon_m\epsilon_p i^{m+p} J_m(A_0[t_0']z) J_p(A_0[t_0'+t]\xi)\right.$$

$$\left.\times \frac{1}{T_0}\int_0^{T_0} dt_0 \cos\omega_0 m t_0 \cos p\omega_0(t_0+t)\right\}, \tag{2.17}$$

where we have used the familiar expansion of $\exp(i\alpha\cos\beta t)$ in a Fourier series,[21] and where $\epsilon_0 = 1$, $\epsilon_m = 2$, $m \geqq 1$. The trigonometric integral has the value $(\cos m\omega_0 t)\delta_m^p/\epsilon_m^p$, so that (2.17) may be written finally

$$F_s(z,\xi;t) = \sum_{m=0}^{\infty}(-1)^m\epsilon_m\cos m\omega_0 t\left\{\frac{1}{T_0'}\int_0^{T_0'} dt_0' J_m(A_0[t_0']z)J_m(A_0[t_0'+t]\xi)\right\}. \tag{2.18}$$

When (2.16) and (2.18) are substituted in the expression for the complete correlation function $R(t)$, we may expand $\exp[-\psi(t)z\xi]$, obtaining

$$R(t) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(-1)^{m+n}\epsilon_m\psi(t)^n\cos m\omega_0 t}{n!}[H_{mn}(t_0')H_{mn}(t_0'+t)]_{\text{av.}}, \tag{2.19}$$

where

$$\left.\begin{aligned} H_{mn}(t_0') &= \frac{1}{2\pi}\int_C z^n f(iz) J_m(zA_0[t_0'])e^{-\psi z^2/2}dz, \\ H_{mn}(t_0'+t) &= \frac{1}{2\pi}\int_{C'}\xi^n f(i\xi)J_m(\xi A_0[t_0'+t])e^{-\psi\xi^2/2}d\xi. \end{aligned}\right\} \tag{2.20}$$

The average indicated above in (2.19) is the time average over the phases of the modulation. The mean power spectrum of the output wave follows from (1.6) and (2.19), and is

$$W(f) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}(-1)^{m+n}\frac{\epsilon_m}{n!}C_{mn}(f), \tag{2.21}$$

where $C_{mn}(f)$ is given by

$$C_{mn}(f) = 4\int_0^{\infty}\psi(t)^n\cos m\omega_0 t\cos\omega t[H_{mn}(t_0')H_{mn}(t_0'+t)]_{\text{av.}}dt, \quad \omega = 2\pi f. \tag{2.22}$$

The quantity $[H_{mn}(t_0')H_{mn}(t_0'+t)]_{\text{av.}}$ is, in fact, a kind of correlation function, by formal comparison with (A2.1), except that here no statistical averages appear because $A_0(t)$ is periodic. We can also write formally

[21] Watson, *Theory of Bessel Functions*, Cambridge Univ. Press, 1944, sec. 2.22.

$$[H_{mn}(t_0')H_{mn}(t_0'+t)]_{\text{av.}}$$

$$= \frac{1}{4\pi^2}\int_C z^n f(iz)e^{-\psi z^2/2}dz \int_{C'} \xi^n f(i\xi)e^{-\psi\xi^2/2}[J_m(zA_0[t_0'])J_m(\xi A_0[t_0'+t])]_{\text{av.}}d\xi$$

$$= \sum_{k=0}^{\infty} \epsilon_k h_{kmn}^2 \cos k\omega_A t, \tag{2.23}$$

where $f_A(=\omega_A/2\pi)$ is the fundamental frequency of $A_0(t)$ and h_{kmn}^2 is a mean-square amplitude, which depends on ψ, the amplitude of $A_0(t)$, and on the dynamic characteristic of the apparatus in question.[22] Equation (2.23) is the result of developing the Bessel functions in series, or the expressions integrated first over z and ξ, and arranging the result as a series in $\cos k\omega_A t$. The explicit evaluation of h_{kmn} is in general a difficult task. In some cases the work is simplified by taking the time average first, in others, by integrating over z and ξ initially. There seems to be no simple way of handling the integration problem presented by (2.23), but usually the average over the phases of the modulation must be performed last. Some illustrative examples are given later in this section.

From (2.23) we observe that (2.22) becomes

$$C_{mn}(f) = \sum_{k=0}^{\infty} \epsilon_k h_{kmn}^2 c_{kmn}(f), \tag{2.24a}$$

where

$$c_{kmn}(f) = 4\int_0^{\infty} \psi(t)^n \cos m\omega_0 t \cos k\omega_A t \cos \omega t dt; \tag{2.24b}$$

the mean output spectrum (2.21) finally takes the form

$$W(f) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{m+n}\frac{\epsilon_m\epsilon_k}{n!}h_{kmn}^2 c_{kmn}(f). \tag{2.25}$$

The effect of the rectifier or similar non-linear device is to "mix" or cross-modulate the noise and signal components with one another so that the (unfiltered) output contains the following three classes of modulation product: (a) *noise*×*noise*, which gives rise to noise, no longer random with Gaussian properties, (b) *noise*×*signal*, which results from the beating of the signal components with the noise wave and which in turn also yields non-Gaussian noise, and finally, (c) *signal*×*signal*, produced by the cross-modulation of the various signal components: this last is entirely periodic and free from noise. A d-c component is often present, but not when the dynamic characteristic is symmetrical, for then the average of the cutput, as well as the input, vanishes. Further subdivision is possible when there is modulation, as we may then distinguish between modulation products generated by the separate (incommensurable) components of the carrier and modulation, but essentially the three main classifications (a)–(c) still hold.

The different contributions to (a)–(c) may be distinguished in (2.25), since for (a), exclusive of the d-c, we may write, with $m=k=0$, $n\geqq 1$,

[22] It is also possible to express the average in terms of the moments $\overline{A_{01}^m A_{02}^p}$, on expanding the Bessel functions with the help of the relation on page 148 of reference 21.

$$W(f)_{\text{noise}\times\text{noise}} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} h_{00n}^2 c_{00n}(f), \tag{2.26a}$$

and for (b) we have $m \geqq 1$, $n \geqq 1$, $k \geqq 1$, giving

$$W(f)_{\text{noise}\times\text{signal}} = 4 \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \frac{(-1)^{m+n}}{n!} h_{kmn}^2 c_{kmn}(f). \tag{2.26b}$$

Type (c) occurs only for $m \geqq 1$, $k \geqq 0$, $n = 0$ or $m \geqq 0$, $k \geqq 1$, $n = 0$:

$$W(f)_{\text{signal}\times\text{signal}} = \sum_{\binom{m=1}{=0}}^{\infty} \sum_{\binom{k=0}{=1}}^{\infty} \epsilon_k (-1)^m h_{km0}^2 \{\delta(f - mf_0 - kf_A) + \delta(f - mf_0 + kf_A)\}, \tag{2.26c}$$

where we have used the delta function to indicate the "discrete" nature of these components, located in frequency at $mf_0 \pm kf_A$. The spectral distributions (a) and (b), on the other hand, representing noise, are continuous. There remains only the d-c, which is specified by $m = n = k = 0$:

$$W_{\text{d-c}} = h_{000}^2 \, 4 \int_0^{\infty} df \left(\int_0^{\infty} \cos \omega t dt \right) = h_{000}^2 \, 2 \int_0^{\infty} \delta(f - 0) df = h_{000}^2. \tag{2.26d}$$

The parts of the correlation function which contribute to (a)–(c) are easily found from (2.19), (2.21), and (1.6).

We turn now to a number of interesting special cases.

I. Unmodulated carrier. This is the first and simplest case involving a signal; A_0 is a constant quantity representing the amplitude of the carrier wave. It is immediately clear that the only non-vanishing terms require $k = 0$ ($m, n \geqq 0$), so that the correlation function reduces to the form given by Eqs. (4.9-7) of reference 2, while $h_{mn} (\equiv h_{mn0})$ is expressed by Eq. (4.9-6). For *noise alone* entering the non linear device, the expressions (2.19)–(2.26d) are still simpler, reducing to the forms given by Eqs. (3.1)–(3.6) of reference 9.

II. Carrier modulated by sine wave.[23] For this important case we have for the signal (2.1)

$$V_s = A_0(1 + \lambda \cos \omega_A t) \cos \omega_0 t, \; A_0(t_0') = A_0(1 + \lambda \cos \omega_A t_0'), \; 0 \leqq \lambda \leqq 1, \tag{2.27}$$

where λ is the modulation index and A_0 is the (peak) amplitude of the carrier component. For the moment no restriction is set on f_A and f_0, save that they be incommensurable, which is another way of stating that $A_0(t)$ and $\cos \omega_0 t$ are uncorrelated. Of course, for $A_0(t)$ to be spoken of as a modulation in the usual sense, the frequency f_A must be much less than f_0.

We require the value of $[H_{mn}(t_0')H_{mn}(t_0' + t)]_{\text{av.}}$. Expansion of the Bessel functions in (2.20) yields various moments of the modulation, viz. $[A_0^{a_1}(t_0')A_0^{a_2}(t_0' + t)]_{\text{av.}}$. The average over t_0' follows in the present instance, *cf.* (2.27), with the aid of the relation

---

[23] For some earlier work on this problem, see J. R. Ragazzini, Proc. I.R.E. **30**, 227, 1942. See also sections 4.1, 4.2, 4.10 of ref. 2. A theoretical discussion of this problem in the case of half-wave rectification has been given by the author in a paper submitted to the I.R.E. A companion paper by Fubini and Johnson verifies the results experimentally.

$$I_{a1,a2}(\phi) = \frac{1}{2\pi}\int_0^{2\pi} (1 + \lambda \cos\theta)^{a_1}(1 + \lambda \cos[\theta + \phi])^{a_2} d\theta, \qquad \phi = \omega_A t,$$

$$= \sum_{j=1}^{a_2} 2^{1-2j}\lambda^{2j}\, {}_{a_2}C_j (-1)^j \sum_{i=0}^{j/2,\,(j-1)/2} {}_jC_i\, {}_j\Delta_i [-a_1]_{j-2i} \frac{(2/\lambda)^{2i}}{(j-2i)!} \tag{2.28}$$

$$\times \cos(j-2i)\phi\, {}_2F_1[\tfrac{1}{2}(j-2i-a_1), \tfrac{1}{2}(j-2i-a_1+1); 1+j-2i; \lambda^2], \tag{2.29}$$

where a_1 and a_2 are integers, the C's are the usual binomial coefficients, and ${}_j\Delta_i = \frac{1}{2}$ when $j = 2i$, and ${}_j\Delta_i = 1, j \neq 2i$. The limits on the second summation apply accordingly as j is even or odd, and $[\alpha]_\beta = \alpha(\alpha+1)\cdots[\alpha+\beta-1]$, $[\alpha]_0 = 1$. The expression (2.29) follows from the development of $(1+\lambda\cos[\theta+\phi])^{a_2}$ in a series in $\cos\theta$ and the result[24]

$$\frac{1}{2\pi}\int_0^{2\pi} (1 + \lambda\cos\theta)^l \cos k\theta\, d\theta$$

$$= \frac{(-l)_k(-\lambda)^k}{2^k k!}\, {}_2F_1[\tfrac{1}{2}(k-l), \tfrac{1}{2}(k-l+1); k+1; \lambda^2]. \tag{2.30}$$

From this it is also easily seen that when $\phi = 0$

$$I_{a1,a2}(0) = {}_2F_1[-\tfrac{1}{2}(a_1+a_2), \tfrac{1}{2}(-a_1-a_2+1); 1; \lambda^2]. \tag{2.31}$$

Equation (2.31) is helpful if one desires to determine the mean power in the wave. A short table of $I_{a_1,a_2}(\phi)$ is given below.

TABLE I

	$a_1=0$	1	2	3	4	5
$a_1=0$	1	1	$1+\lambda^2/2$	$1+3\lambda^2/2$	$1+3\lambda^2/2+3\lambda^4/8$	$1+5\lambda^2+15\lambda^4/8$
1	1	$1+(\lambda^2/2)\cos\phi$	$1+\lambda^2(\frac{1}{2}+\cos\phi)$	$1+(3\lambda^2/2)(1+\cos\phi)+(3\lambda^4/8)\cos\phi$	$1+\lambda^2(3+2\cos\phi)+(3\lambda^4/8)(1+4\cos\phi)$	
2	$1+\lambda^2/2$	$1+\lambda^2(\frac{1}{2}+\cos\phi)$	$1+\lambda^2(1+2\cos\phi)+(\lambda^4/8)(2+\cos 2\phi)$	$1+\lambda^2(2+3\cos\phi)+(3\lambda^4/8)(2+2\cos\phi+\cos 2\phi)$		
3	$1+3\lambda^2/2$	$1+(3\lambda^2/2)(1+\cos\phi)+(3\lambda^4/8)\cos\phi$	$1+\lambda^2(2+3\cos\phi)+(3\lambda^4/8)(2+2\cos\phi+\cos 2\phi)$			
4	$1+3\lambda^2/2+3\lambda^4/8$	$1+\lambda^2(3+2\cos\phi)+(3\lambda^4/8)(1+4\cos\phi)$				
5	$1+5\lambda^2+15\lambda^4/8$					

Once the correlation function has been found it is comparatively easy to determine the mean power in the output by setting $t = 0$ in the appropriate expressions for $R(t)$. As shown by (A2.6) and as discussed more fully in Appendix II, this is equivalent to integrating over all spectral components, "discrete" or continuous. It is also evident that the *mean total power* W_r in the output wave *is independent of spectral shape* of the input distribution provided all significant components are included in our calculation. Thus from (2.19) and (2.23) we may write, setting $t = 0$,

[24] Reference 2, Eq. (4.2.17).

$$W_r = R(0) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(-1)^{m+n}}{n!}\epsilon_m\psi(0)^n[H_{mn}(t_0')^2]_{av.}$$

$$= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\frac{(-1)^{m+n}\epsilon_m\epsilon_k\psi(0)^n}{n!}\,h_{kmn}^2 \tag{2.32}$$

for a general, modulated carrier rectified in the presence of random noise. The power in the continuum W_r is readily obtained from (2.32) and (2.26d), viz., $W_r = W_r - W_{d-c} = W - h_{000}^2$, or by omitting from the sum the term in $k=m=n=0$ in (2.32). In a similar fashion, to determine the portion of the total power output attributable to modulation products formed by (a), *noise*×*noise*, (b), *noise*×*signal*, and (c), *signal*×*signal*, we have only to consider in (2.32) those terms for which $m=k=0$, $n\geqq 1$, $m\geqq 1$, $n\geqq 1$, $k\geqq 1$, and $m\geqq 1$, $k\geqq 0$, $n=0$ or $m\geqq 0$, $k\geqq 1$, $n=0$, respectively.

One may also calculate the mean power in the output without knowledge of the correlation function, for if $W_1(V_{in})$ is the first order probability density for the input wave (2.10), then by definition the mean total power dissipated by the rectified disturbance in a unit resistance is

$$W_r = \overline{g(V_{in})^2} = \int_{-\infty}^{\infty} g(V_{in})^2 W_1(V_{in})dV_{in}, \tag{2.33a}$$

with a corresponding expression for the d-c component:

$$W_{d\text{-}c} = [\overline{g(V_{in})}]^2 = \left\{\int_{-\infty}^{\infty} g(V_{in})W_1(V_{in})dV_{in}\right\}^2. \tag{2.33b}$$

The quantity $W_1(V_{in})$ has been derived independently by Rice[25] and Bennett[26] for the case of the simple sine wave and noise. For noise alone it is given by (2.5); the distribution is more complicated when a signal is present. The drawback to (2.33a) as a method of determining the power output, however, lies in the difficulty of performing the integrations when g is discontinuous, for instance in the half-wave linear rectifier, where $g>0$, $V_{in}>0$; $g=0$, $V<0$. Equation (2.33a) is useful only when the characteristic g exhibits no abrupt variations, cf., "small-signal" detection, in which the incoming disturbance has a mean square amplitude small compared with the curvature of the dynamic path (see Sec. 8), or when noise alone is subject to rectification, see Ref. 9. In such cases (2.33a) is to be preferred because it yields results in terms of a finite number of (tabulated) functions, whereas the corresponding expression for W_r obtained from $R(0)$ is given as an infinite series which does not converge too rapidly. On the other hand, use of the correlation function has the advantage that from it can be predicted the order and nature of particular modulation products or groups of products, i.e., whether the contribution arises from *noise*×*noise* or *noise*×*signal*, etc. This, as we shall see in the next section, is an important property of the correlation function, especially useful when the incoming wave is spectrally narrow compared with its central frequency, for then the output is composed of separate harmonic "zones," any one of which may in principle be isolated with a suitable filter and examined.

3. Narrow band waves. So far, in discussing spectra and correlation functions no restrictions have been placed on the spectral nature of the input wave, which is com-

[25] Reference 2, Eq. (3.10.6).
[26] Reference 6, Eq. (3.6).

posed, in its general form, of periodic or of a set of periodic components and of a noise term. Moreover, in determining the mean power spectrum of the output disturbance, no additional modification of the wave has been assumed other than that attributable to the non linear $i_p - e_g$ (output current vs. applied voltage) path of the rectifying device. In somewhat different language this means that the circuits of our apparatus contain only filters whose amplitude response is uniform (for all frequencies) and whose phase shifts vary in linear fashion with the frequency. This means that we restrict ourselves to non linear elements whose instantaneous output does not affect the time dependence of the incoming wave, i.e., to resistive loads, no feedback, or at worst to plate (and grid) circuits whose reactive components are small compared with their resistive ones—all of which is required for essentially one-valued dynamic paths (Fig. 1)—then we may introduce the effect of filters into the analysis, observing that now they simply modify the *frequency* spectrum. In the case of filtering the input the spectral shape is altered, but the quality of the wave is not altered; it is still a Gaussian random process. On the other hand, if the output is filtered, not only may the original, i.e., the unfiltered output, be changed spectrally, but also the quality of

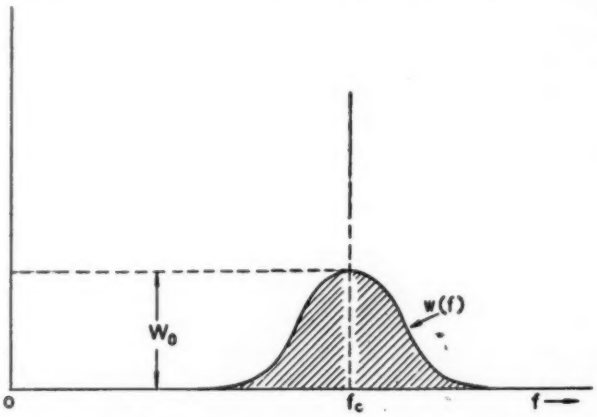


FIG. 2. Narrow-band input noise (power) spectrum, centered about f_c.

the wave is affected, in as much as the output is no longer Gaussian, because of the intervening non linear operation. The distributions of the output amplitude, however, remain unchanged by a filter, *provided that the pass-band of this filter is sufficiently wide compared with the spectrum of the transmitted wave.* If it is not, then there is a tendency of the filter to restore randomness to the distribution by filtering out the higher frequencies characteristic of the distortions due to curvature of the dynamic path and to the "ceiling" produced by cut-off, saturation, or both.* But we can still determine the mean power and mean power spectrum, although the distributions $W_1, W_2, \cdots, W_n$ are at present beyond our powers, by observing that

$$W(f)_{\text{filter in output}} = W(f)_{\text{inf. filter in output}} \cdot |G(f)|^2,$$

where $|G(f)|$ is the modulus of the output filter response.

* *Note added in proof:* The difficult problem of determining the distribution of the amplitudes after rectification and *arbitrary* filtering has been solved by M. Kac and A. J. F. Siegert, J. Appl. Phys., **18,** 383 (1947) for a quadratic detector.

When the spectral width, measured between points at which the mean power spectrum is but a few percent of its maximum value, is small compared to the central or "resonant" frequency, we term the wave narrow-band. A precise definition, of course, requires a knowledge of the spectral shape, whether it has one or more maxima, where we measure width, etc., but here we shall assume that the spectrum is a simple, symmetrical distribution about a single maximum, as shown in Fig. 2. Now an important property of narrow band noise waves is that their correlation functions may be expressed as the product of the correlation functions of a slowly and of a rapidly varying term, corresponding to an envelope or "modulation," and to a "carrier," having the central or resonant frequency f_c of the disturbance. (This is shown to hold for symmetrical noise spectra in Appendix II (A2.8); for a signal the same demonstration may be made, except now, of course, we have a series of discrete components distributed about a true carrier, rather than a continuum about a central frequency.) The correlation function and the spectrum of the output follow directly from (2.14) and (1.6) as before, but instead of expanding $\exp\left[-\psi(t)z\xi\right]$ at once to obtain the form (2.19), we may use this property of the input correlation, viz:

$$\psi(t) = \psi(0)r_i(t)\cos\omega_c t \tag{A2.8}$$

to show that the outgoing wave is composed of an infinite number of spectral bands centered about harmonics of the central or of the carrier frequency, or of their modulation products. Generally speaking, these bands will overlap to varying degrees, if f_0 and f_c are different; if $f_0=f_c$, which is the important case in practice, the modulation products involving f_0 and f_c will coincide exactly. The different behaviors are apparent when we attempt to examine a given spectral region, say one centered about the carrier f_0. In the former instance ($f_0\neq f_c$) our filter* will exclude part of the spectral zone due to the one and include some of the zone associated with the other, while for the latter ($f_c=f_0$), our filter passes contributions symmetrically disposed about f_0. For the harmonic regions of higher spectral order, i.e., lf_0, lf_c, $l\geqq 2$, the former effect becomes more pronounced, since now the maxima of the distributions due to the noise and signal are separated by approximately $l|f_0-f_c|$ cycles. Further there is a "smearing-out" of the spectra, because all orders of *noise*×*noise*, *noise*×*signal*, etc., components are generated. This accounts for the broadening and filling-in of the various output spectra; smearing of the carrier and modulation is also observed. When $l=0$, we obtain the low-frequency output, or envelope, distorted or not depending on the nature of the tube's dynamic path. This spectral region is usually of principal interest in reception, while in transmission the zone $l=1$ is of chief concern, where now our rectifier assumes the rôle of mixer (though with a different characteristic, of course). Figure 3 illustrates this discussion.

Now with the help of (A2.8) and the expansion

$$e^{-\psi r_0(t)\cos\omega_0 t} = \sum_{p=0}^{\infty} \epsilon_p(-1)^p I_p(z\xi\psi r_0(t))\cos p\omega_c t \tag{3.1}$$

we may express the correlation function (2.14) as

* For this purpose all our filters are assumed, to be ideal, i.e., to have a uniform response over a frequency range wide enough to include (almost) all the components in a given zone. We say almost, because the non linear device produces some frequencies which will lie outside the finite pass-band (see the end of Appendix II).

$$R(t) = \sum_{m=0}^{\infty}\sum_{p=0}^{\infty} \epsilon_m \epsilon_p (-1)^{m+p} \cos p\omega_c t \cos m\omega_0 t \left\{ \frac{1}{4\pi^2} \int_C f(iz) J_m(zA_0[t_0']) e^{-\psi z^2/2} dz \right.$$
$$\left. \times \int_{C'} f(i\xi) J_m(\xi A_0[t_0' + t]) e^{-\psi \xi^2/2} d\xi \cdot I_p(z\xi r\psi) \right\}_{\text{av.}} . \tag{3.2}$$

Equation (3.2) is general for all f_0 and f_c. However, in the case of principal interest carrier and central frequency coincide, so that $f_0 = f_c$, and (3.2) reduces to a series of

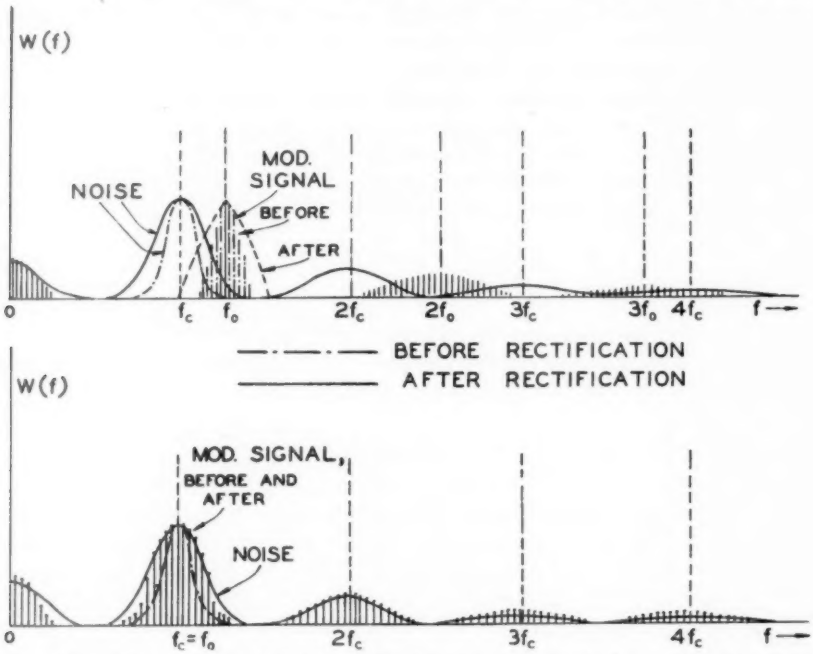


FIG. 3. Narrow-band noise and modulated signal spectra, before and after rectification. The upper of the two applies when $f_c \neq f_0$, where the spectral width is much less than either f_c or f_0. The lower illustrates the spectra when $f_c = f_0$. The normalization is arbitrary.

distinct harmonic regions distributed about lf_0, $l = 0, 1, 2, 3 \cdots$, the harmonics of the carrier. Thus, from the fact that replacing $\sum_{m=0}^{\infty}\sum_{p=0}^{\infty} \epsilon_m \epsilon_p (\ )$ by $\sum_{m=-\infty}^{\infty}\sum_{p=-\infty}^{\infty} (\ )$, since $J_m(\alpha)J_m(\beta) = J_{-m}(\alpha)J_{-m}(\beta)$ and $I_p(\gamma) = I_{-p}(\gamma)$, and collecting all terms of $\cos p\omega_0 t \cos m\omega_0 t$ which yield $\cos l\omega_0 t$, we observe that the correlation function for the lth band is

$$R_l(t) = \frac{(-1)^l \cos l\omega_0 t}{4\pi^2} \sum_{m=0}^{\infty} \epsilon_{|m-l|} \left\{ \int_C f(iz) J_m(A_0(t_0')z) e^{-\psi z^2/2} dz \right.$$
$$\left. \times \int_{C'} f(i\xi) J_m(A_0[t_0' + t]\xi) I_{|m-l|}(\psi \cdot z\xi r_0) e^{-\psi \xi^2/2} d\xi \right\}_{\text{av.}}$$
$$= (-1)^l \cos l\omega_0 t \sum_{m=0}^{\infty} \frac{\epsilon_{|m-l|} \psi^{|m-l|}}{2^{|m-l|}} \sum_{q=0}^{\infty} \frac{\psi^{2q} r_0(t)^{2q+|m-l|}}{2^{2q} q! (q + |m-l|)!}$$
$$\times [H_{m,|m-l|+2q}(t_0') \cdot H_{n,|m-l|+2q}(t_0' + t)]_{\text{av.}}, \tag{3.3}$$

on expanding the Bessel function $I_{|m-l|}$. The complete correlation function is, of course, $\sum_{l=0}^{\infty} R_l(t)$. By (A2.6) and (2.23) the power in any zone is

$$R_l(0) = (-1)^l \sum_{m=0}^{\infty} \frac{\psi^{|m-l|}\epsilon_{|m-l|}}{2^{|m-l|}} \sum_{q=0}^{\infty} \frac{\psi^{2q}2^{-2q}}{q!(q+|m-l|)!} \sum_{k=0}^{\infty} \epsilon_k h^2_{k,m,|m-l+2q}, \tag{3.4}$$

since $r_0(0) = 1$. Here also we see an illustration of the theorem stated previously, cf. (2.32) and Appendix II, namely, that the power in a spectral region, here an harmonic zone rather than the entire output spectrum, is independent of the spectral shape of the input noise and also of the signal spectrum. This naturally simplifies the calculation of the power, once $R_l(t)$ has been found.

4. The biased νth-law rectifier. In the preceding section certain general aspects of the theory of modulated signals and noise passed through non linear devices have been outlined. Specific results in each instance depend on the evaluation of integrals of the form (2.20). Here attention is directed to the important case of the biased νth-law detector, whose dynamic characteristic $g(V)$ is given by

$$\begin{aligned} I = g(V) &= \beta(V - b_0)^\nu, \qquad V > b_0, \qquad \nu > 0 \\ &= 0, \qquad\qquad V < b_0, \end{aligned} \tag{4.1}$$

and β has suitable dimensions for the output to be a current when the incoming wave is a voltage; b_0 represents the cut-off voltage, measured from the operating point in the manner of Fig. 1. The Fourier transform of g is readily shown to be

$$f(iz) = \int_0^\infty e^{-(iz)V} g(V)dV = \frac{\beta\Gamma(\nu+1)e^{-izb_0}}{(iz)^{\nu+1}}. \tag{4.2}$$

With the help of (4.2) the integrals (2.20) may now be determined, in a variety of ways leading to different, but equivalent results, the form of either depending on whether $\exp(-ib_0 z)$ or $J_m(A_0 z)$ is expanded in a series, followed in each instance by termwise integration. The former approach yields a series in $b \equiv b_0/\psi^{1/2}$, where the coefficients of the various powers of b are themselves series in the parameter $p(t) \equiv [A_0(t)]^2/2\psi$, while the latter is expressible as a series in $p(t)$, the coefficient of whose general term likewise is an expansion in b. The two developments, while valid for all values of the parameters b, $p(t)$, and ψ, vary in their usefulness for different ranges of b, $p(t)$, and ψ.

Let us consider first the case where $\exp(-izb_0)$ is given as a series. The general integral (2.20) becomes accordingly, with the aid of (4.2),

$$H_{mn}(t) = \frac{\beta\Gamma(\nu+1)}{2\pi} \int_C z^n e^{-izb_0} J_m(zA_0(t)) e^{-\psi z^2/2} dz/(iz)^{\nu+1} \tag{4.3}$$

$$= \frac{\beta\Gamma(\nu+1)}{2\pi} i^{-\nu-1} \sum_{k=0}^{\infty} \frac{(-i)^k b_0^k}{k!} \int_C z^{k+n+\nu-1} J_m(zA_0(t)) e^{-\psi z^2/2} dz$$

$$= \frac{\beta\Gamma(\nu+1)e^{-(m+n)\pi i/2}}{2m!} \left(\frac{\psi}{2}\right)^{(\nu-n)/2} p(t)^{m/2}$$

$$\cdot \sum_{k=0}^{\infty} \frac{(-1)^k 2^{k/2} b^k \, {}_1F_1[(k+m+n-\nu)/2; m+1; -p(t)]}{k!\Gamma[(2+\nu-k-m-n)/2]}, \tag{4.4}$$

this last from Appendix III, Eq. (A3.15), where also are mentioned some of the prop-

erties of the confluent hypergeometric function ${}_1F_1$. For an unmodulated carrier $p(t)$ is merely a constant $p=A_0^2/2\psi$. Observe also that when $(2+\nu-k-m-n)$ is zero or an even *negative* integer the modulation products contributed by such terms vanish, in virtue of the poles of $\Gamma[(2+\nu-k-m-n)/2]$. Further, when ν is integral, only odd values of k appear, $m+n$ even, and only even values of k when $m+n$ is odd (except for $2+\nu>k+m+n>0$). Equation (4.4) is best suited for computation when $b=b_0/\psi^{1/2}$ is of the order of unity or less, even for a wide range of values of $p(t)$. Values of $b>1$ yield too slow a convergence, and the alternative development is then needed.

The expansion of the Bessel function in (4.3) gives us

$$H_{mn}(t) = \frac{\beta\Gamma(\nu+1)}{2\pi} i^{-\nu-1}\sum_{k=0}^{\infty} \frac{(-1)^k A_0(t)^{2k+m}}{k!(k+m)!2^{2k+m}}\int_C z^{2k+m+n-\nu-1}e^{-ib_0z-\psi z^2/2}dz,$$

and from Eq. (A3.17) we obtain finally

$$H_{mn}(t) = \frac{\beta\Gamma(\nu+1)}{2} e^{-\pi i(m+n)/2}\left(\frac{\psi}{2}\right)^{(\nu-n)/2} p(t)^{m/2}\sum_{k=0}^{\infty}\alpha_{mnk}p(t)^k, \tag{4.5}$$

where

$$\alpha_{mnk} = \frac{1}{k!(k+m)!}\left\{\frac{{}_1F_1[(2k+m+n-\nu)/2;\,1/2;\,-b^2/2]}{\Gamma[(2+\nu-2k-m-n)/2]}\right.$$
$$\left. -\sqrt{2}\,b\,\frac{{}_1F_1[(2k+m+n-\nu-1)/2;\,3/2;\,-b^2/2]}{\Gamma[(1+\nu-2k-m-n)/2]}\right\}. \tag{4.6}$$

Here large values of b are clearly more easily handled, especially when the asymptotic series for ${}_1F_1$, cf. (A3.3), can be used. Also equations (4.5) and (4.6) in general offer a more satisfactory form from which to determine the time average over the phases of the modulation, viz:

$$[H_{mn}(t_0')H_{mn}(t_0'+t)]_{\text{av.}}$$
$$= \frac{\beta^2\Gamma(\nu+1)^2(-1)^{m+n}}{4}\left(\frac{\psi}{2}\right)^{\nu-n}\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\alpha_{mnk_1}\alpha_{mnk_2}[p(t_0')^{m/2+k_1}p(t_0'+t)^{m/2+k_2}]_{\text{av.}}, \tag{4.7}$$

when the spectrum of the output is desired, cf. Eqs. (2.21), (2.22), since a double series is sufficient, whereas if (4.4) were used, a fourfold infinite series would result for the average. A similar superiority of (4.7) over (4.4) is noted in the limiting case of $p(t)\ll 1$. Observe also from (A3.9) that for integral values of $\nu(>0)$ the various α_{mnk} reduce to the functions defined in Appendix III (A3.4), and are thus expressible as Hermitian polynomials.

A detailed discussion of the above in the case of the biased linear rectifier or mixer is reserved for a later paper.

Limiting cases are of interest: We consider first:

Case I: $b_0\to 0$, $A_0(t), \psi\neq 0$ *and finite.* Equations (4.4) and (4.7) reduce, equivalently, to

$$H_{mn}(t)_{b_0\to 0} = \frac{\beta\Gamma(\gamma+1)}{2m!} e^{-(m+n)\pi i/2}\left(\frac{\psi}{2}\right)^{(\nu-m)/2} p(t)^{m/2}\left\{\frac{{}_1F_1[(n+m-\nu)/2;\,m+1;\,-p(t)]}{\Gamma[(2+\nu-m-n)/2]}\right.$$
$$\left. -b\sqrt{2}\,\frac{{}_1F_1[(1+m+n-\nu)/2;\,m+1;\,-p(t)]}{\Gamma[(1+\nu-m-n)/2]}+\cdots\right\}. \tag{4.8}$$

From (4.8) the expressions for half-wave rectification, $b=0$, follow immediately.

Case II: $b_0 \to \pm\infty$, $A_0(t), \psi$ *finite.* Here it is convenient to use the second form (4.5) of $H_{mn}(t)$ and apply the asymptotic series (A3.3) for ${}_1F_1$. It then may be shown that

$$H_{mn}(t)_{b_0 \to \infty} \cong 0. \tag{4.9}$$

since $\alpha_{mnk} \to 0$ for *all* terms in the asymptotic development. This is to be expected physically, since with an indefinitely large cut-off voltage the finite signal amplitudes and even the possibly infinite noise peaks are not transmitted, the latter because they arise a vanishingly small fraction of the time, since the mean power in the input signal and noise waves is finite. On the other hand, when $b_0 \to -\infty$, we find from (A3.3) and (4.5) that

$$H_{mn}(t)_{b_0 \to -\infty} \cong \frac{\beta\Gamma(\nu+1)e^{-(m+n)\pi i/2}\sqrt{\pi}\,2^{n-\nu}\,|\,b_0\,|^{-m-n+\nu}A_0(t)^m}{m!\Gamma[(1+\nu-m-n)/2]\Gamma[(2+\nu-m-n)/2]} \times \left\{1+\frac{\psi}{2b_0^2}\left[1-\frac{A_0^2(t)}{2\psi(m+1)}\right](m+n-\nu+1)(m+n-\nu)+\cdots\right\}, \tag{4.10}$$

which shows that for $\nu > m+n$ the output contains increasingly large terms, while for $\nu < m+n$ all such contributions become negligibly small. As an example, consider the linear rectifier, $\nu=1$. Here only the terms $(m, n)=(0, 0), (0, 1), (1, 0)$ are significant:

$$H_{00} = \beta\,|\,b_0\,|; \quad H_{01} = -i\beta; \quad H_{10}(t) = -i\beta A_0(t)/2; \quad H_{mn}(t) \to 0,\ m+n>1. \tag{4.10a}$$

The first relation becomes infinite, since infinite (negative) cut-off voltage must be supplied when $b_0 \to -\infty$; the second result, when substituted in (2.21) and (2.22), gives the mean power spectrum of the output noise, which is easily seen to be the same as the input spectrum except for the constant factor β, and finally, the third expression is observed to be the mean power spectrum of the input signal, undistorted because now the linear rectifier is essentially a linear amplifier. In general (4.10) may be said to hold for a small-signal νth-law detector (see Sec. 7).

Case III: $A_0(t) \to 0$, b_0, ψ *finite.* Here the modulation or carrier contribution is allowed to become small. Then the principal term in $H_{mn}(t)$ obtains when $m=0$, and correction terms when $m=0, k=1$, $m=1, k=0$, and $m=2, k=0$. Either (4.4) or (4.5) reduces to the following

$$\begin{aligned} H_{0n}(t)_{p(t)\to 0} &= \frac{\beta\Gamma(\gamma+1)e^{-\pi ni/2}}{2}\left(\frac{\psi}{2}\right)^{(\nu-n)/2}\{\alpha_{0n0}+\alpha_{0n1}p(t)+\cdots\}, \\ H_{1n}(t) &= -\frac{i\beta\Gamma(\nu+1)e^{-\pi ni/2}}{2}\left(\frac{\psi}{2}\right)^{(\nu-n)/2}\{p(t)^{1/2}\alpha_{1n0}+\cdots\}, \\ H_{2n}(t) &= -\frac{\beta\Gamma(\gamma+1)e^{-\pi ni/2}}{2}\left(\frac{\psi}{2}\right)^{(\nu-n)/2}\{p(t)\alpha_{2n0}+\cdots\}. \end{aligned} \tag{4.11}$$

For noise alone we have

$$H_{0n}(t) = h_{0n} = \frac{\beta\Gamma(\nu+1)e^{-\pi ni/2}}{2}\left(\frac{\psi}{2}\right)^{(\nu-n)/2}\left\{\frac{{}_1F_1[(n-\nu)/2;\,1/2;\,-b^2/2]}{\Gamma[(2+\nu-n)/2]}-\right.$$

$$- b\sqrt{2}\,\frac{{}_1F_1[(n-\nu+1)/2;\,3/2;\,-b^2/2]}{\Gamma[(1+\nu-n)/2]}\Big\}, \quad (4.12)$$

and in the special cases $\nu=1$ and $\nu=2$ we obtain with the aid of (A3.9) the equations of Reference 9 when there is no saturation.

Case IV: $A_0(t)\to\infty$, b_0, ψ *finite*. The signal is taken to be very large, and again the asymptotic relation (A3.3) is in order, this time applied to (4.4) as $p(t)\to\infty$. The result is

$$H_{mn}(t)_{A_0(t)\to\infty} \leqq \frac{\beta\Gamma(\nu+1)(-1)^{(m+n)/2}2^{n-\nu-1}A_0(t)^{\nu-n}}{2\Gamma[(2+\nu-n+m)/2]}\Big\{\frac{1}{\Gamma[(2+\nu-n-m)/2]}$$
$$-\frac{2b_0\Gamma[(2+\nu-n+m)/2]}{A_0(t)\Gamma[(1+\nu-n+m)/2](1+\nu-n-m)/2}+\cdots\Big\}, \quad (4.13)$$

and including terms in $A_0(t)^{-1}$ we may write

$$H_{m0}(t) \leqq \beta\Gamma(\nu+1)2^{-\nu-1}(-i)^m A_0(t)^\nu\Big\{\frac{1}{\Gamma[(2+\nu+m)/2]\Gamma[(2+\nu-m)/2]}$$
$$-\frac{2b_0\Gamma(2+\nu+m)}{A_0(t)\Gamma[(1+\nu+m)/2]\Gamma[(1+\nu+m/2)]}+\cdots\Big\},$$
$$H_{m1}(t) \leqq i\beta\Gamma(\nu+1)2^{-\nu}(-i)^m A_0(t)^\nu$$
$$\times\Big\{\frac{1}{\Gamma[(1+\nu+m)/2]\Gamma[(1+\nu-m)/2]A_0(t)}+\cdots\Big\}, \text{ etc.} \quad (4.14)$$

This shows the complete suppression of the noise when $A_0(t)\gg\psi^{1/2}$. Further, since the cut-off and r-m-s noise voltages are comparable in that they are far smaller than the signal, they, too, are negligible in their effect, and we have essentially half-wave rectification of a modulated signal alone. In the special instance of the low-frequency output of the half-wave linear detector, we have from (4.14), $m=0$, $H_{00}(t)\leqq\beta A_0(t)/\pi$, in agreement with Eq. (40) of reference 8.

Case V: $\psi\to\infty$, b_0, $A_0(t)$ *finite*. This corresponds to the case in which the noise overwhelms the signal and much exceeds the cut-off voltage. Equations (4.4) or (4.5) reduce to

$$H_{mn}(t)_{\psi\to\infty} \leqq \frac{\beta\Gamma(\nu+1)}{2m!}(-1)^{(m+n)/2}\psi^{(\nu-n-m)/2}2^{(n-\nu-m)/2}A_0(t)^m$$
$$\cdot\Big\{\frac{{}_1F_1[(n+m-\nu)/2;\,m+1;\,-p(t)]}{\Gamma[(2+\nu-m-n)/2]}$$
$$-\frac{b_0\sqrt{2}}{\psi^{1/2}}\,\frac{{}_1F_1[(n+m+1-\nu)/2;\,m+1;\,-p(t)]}{\Gamma[(1+\gamma-m-n)/2]}$$
$$+\frac{b_0^2}{\psi}\,\frac{{}_1F_1[(m+n+2-\nu)/2;\,m+1;\,-p(t)]}{\Gamma[(\nu-m-n)/2]}+\cdots\Big\}, \quad (4.15)$$

and the significant terms are

$$H_{00}(t) \cong \frac{\beta\Gamma(\nu+1)\psi^{\nu/2}2^{-\nu/2}}{2\Gamma[(2+\nu)/2]}\left\{1+\frac{b_0\sqrt{2}\,\Gamma[(2+\nu)/2]}{\psi^{1/2}\Gamma[(1+\nu)/2]}\right.$$
$$\left.+\frac{1}{\psi}\left[\frac{\nu A_0^2(t)}{4}+\frac{b_0^2\Gamma[(2+\nu)/2]}{\Gamma(\nu/2)}\right]+\cdots\right\},$$

$$H_{01}(t) \cong -\frac{i\beta\Gamma(\nu+1)\psi^{(\nu-1)/2}2^{-(\nu+1)/2}}{\Gamma[(\nu+1)/2]}\left\{1-\frac{b_0\sqrt{2}}{\psi^{1/2}}\,\frac{\Gamma[(1+\nu)/2]}{\Gamma(\nu/2)}+\cdots\right\},$$

$$H_{10}(t) \cong -\frac{i\beta\Gamma(\nu+1)\psi^{\nu/2}2^{-(\nu+3)/2}}{\Gamma[(1+\nu)/2]}\left\{\psi^{-1/2}-\frac{b_0\sqrt{2}}{\psi}\,\frac{\Gamma[(1+\nu)/2]}{\Gamma(\nu/2)}+\cdots\right\},$$

$$\vdots \qquad\qquad \vdots$$

$$H_{02}(t) \cong -\beta\Gamma(\nu+1)\psi^{(\nu/2)-1}2^{-\nu/2}/\Gamma(\nu/2);$$

$$H_{20}(t) \cong -\beta\Gamma(\nu+1)\psi^{\nu/2}\cdot 2^{-(\nu/2)-3}A_0(t)^2/\psi\Gamma(\nu/2), \text{ etc.} \tag{4.16}$$

Observe that the terms containing pure-signal or cross-term contributions, $n=0$, or $m,\ n\neq 0$, vanish at least as $\psi^{-1/2}$, and only noise is left. Thus when the input noise voltage becomes sufficiently great in comparison with the signal, the latter is surpressed. In particular, for the half-wave linear and quadratic rectifier, the low-frequency amplitude of the output ($m=n=0$) becomes from (4.16)

$$\left.\begin{aligned} H_{00}(t)-(d-c) &= \frac{\beta\Gamma(\nu+1)\psi^{\nu/2}}{2^{(\nu/2)+3}\Gamma[(\nu+2)/2]}\left(\frac{\nu A_0^2}{\psi}\right)=\frac{\beta\psi^{1/2}}{2\sqrt{2\pi}}\,[A_0^2(t)/2\psi]_{\nu=1,b_0=0},\ p(t)\ll 1 \\ &= \beta[A_0^2(t)/2]_{\nu=2,b_0=0}, \end{aligned}\right\} \tag{4.17}$$

the first of which is Eq. (39) of reference 8, and the second, Eq. (29).

Case VI: $\psi\to 0$, b_0, $A_0(t)$ *finite.* Here it is convenient to distinguish two cases in the limit, one where $|b_0|>A_0(t)\geqq 0$ and the other when $A_0(t)>|b_0\geqq 0$ For the first we use (4.5) and (4.6) as $b\to\infty$, along with (A3.3) to obtain the not unexpected result that

$$H_{nn}(t)_{\psi\to 0} \cong 0, \qquad b_0 > A_0(t) \geqq 0, \tag{4.18}$$

since the cut-off voltage exceeds the input signal (envelope), so that the wave is not passed. However, when $|b_0|>A_0(t)$, $b_0<0$, we have a different situation, where now the incoming disturbance, which is essentially pure signal (hence $n=0$ in the limit $\psi=0$), is transmitted without distortion due to cut-off or saturation effects, albeit with distortion due to the non linear nature of the dynamic path, $\nu\neq 1$. The expression for the amplitude is accordingly

$$H_{m0}(t) = \frac{\beta\Gamma(\nu+1)i^{-\nu-1}}{2\pi}\int_C e^{-izb_0}J_m(A_0(t)z)\,dz/dz^{\nu+1}. \tag{4.19}$$

When $b_0<0$, $|b_0|>A_0(t)$ we may use (A3.8) in conjunction with (A3.19) to get finally

$$H_{m0}(t) = \frac{\beta\Gamma(\nu+1)e^{-\pi mi/2}\sqrt{\pi}\,|\,b_0\,|^2}{2^\nu\Gamma[(2+\nu-m)/2]\Gamma[(1+\nu-m)/2]m!}\left[\frac{A_0(t)}{|\,b_0\,|}\right]^m {}_2F_1(m-\nu,\, m-\nu+1;$$

$$m+1;\, A_0^2(t)/b_0^2), \qquad b_0<0,\ |\,b_0\,|>A_0(t),\ \psi=0. \tag{4.20}$$

On the other hand, when $|\,b_0\,|<A_0(t)$ we obtain from (4.4), including a correction term,

$$H_{mn}(t)_{\psi\to 0} = \beta\Gamma(\nu+1)e^{-(m+n)\pi i/2}2^{n-\nu-1}A_0(t)^{\nu-n}$$

$$\cdot\sum_{k=0}^{\infty}\frac{(-1)^k 2^k b_0^k\left\{1+\psi(m+n+k-\nu)(n-m+k-\nu)/2A_0(t)^2}{k!A_0(t)^k\Gamma[(2+\nu-k-m-n)/2]\Gamma[(2+\nu-k+m-n)/2]}+\cdots\right\}, \tag{4.21}$$

and when $\psi=0$ we may sum (4.21), $n=0$, or use (A3.19) to get

$$H_{m0}(t)_{\psi=0} = \frac{\beta\Gamma(\nu+1)}{2^{\nu+1}}e^{-m\pi i/2}A_0(t)^\nu\left\{\frac{{}_2F_1[(m-\nu)/2,\,(-m-\nu)/2;\,\frac{1}{2};\,b_0^2/A_0(t)^2]}{\Gamma[(2+\nu+m)/2]\Gamma[(2+\nu-m)/2]}\right.$$

$$\left.-\frac{2b_0}{A_0(t)}\,\frac{{}_2F_1[(m-\nu+1)/2,\,(-m-\nu+1)/2;\,\frac{3}{2};\,b_0^2/A_0(t)^2]}{\Gamma[(1+\nu+m)/2]\Gamma[(1+\nu-m)/2]}\right\}, \tag{4.22}$$

$$0\leqq|\,b_0\,|<A_0(t).$$

For half-wave detection we have the simple result

$$H_{m0}(t)_{b_0=0,\psi=0} = \frac{\beta\Gamma(\nu+1)}{2^{\nu+1}}\cdot\frac{e^{-m\pi i/2}A_0(t)^\nu}{\Gamma[(2+\nu+m)/2]\Gamma[(2+\nu-m)/2]}\cdot \tag{4.23}$$

Power and spectra may be obtained in the usual manner (see Sec. 2).

5. Probability density of the envelope and phase of a modulated signal and noise. In handling the problem of passage of noise or a signal and noise through a non linear device, it is sometimes convenient first to determine the various first and second order probability densities associated with the incoming wave, and then with their help derive the expressions for the mean power output and the correlation function associated with the transmitted disturbance, after it has been modified by a rectifier or similar non linear apparatus, cf. Eq. (1.7b), for example. Now in particular when the noise and signal are narrow-band, *vide* Sec. 3, this method, alternative to the use of the characteristic function [Eqs. (2.10)–(2.16)] suggests itself. The purpose of this section and of section 6 following is to obtain explicit expressions for the probability density of the envelope and phase of the general type of modulated carrier in the presence of narrow-band noise. Only in such circumstances may one properly speak of an envelope or of phase—i.e., when the part of the incoming wave due to noise and modulation is essentially slowly varying in comparison with the carrier frequency f_0 and the central frequency f_c of the noise band. After rectification it is usually this envelope, the low-frequency part of the disturbance, that is observed.

The input is

$$V(t) = A_0(t)\cos\omega_0 t + V(t)_N$$

$$= (A\cos\omega_d t + V_c)\cos\omega_c t + (V_s - A\sin\omega_d t)\sin\omega_c t, \tag{5.1}$$

where f_d is the difference frequency $f_d = f_c - f_0$ and $A = A_0(t)$, from (2.1) and (2.2). Here V_c is the component of the noise "in phase" with $\cos \omega_c t$ and V_s is the component "in phase" with $\sin \omega_c t$, e.g.:

$$\begin{aligned} V_c &= \sum_{n=1}^{\infty} (a_n \cos \omega_n' t + b_n \sin \omega_n' t), \\ V_s &= \sum_{n=1}^{\infty} (-a_n \sin \omega_n' t + b_n \cos \omega_n' t); \qquad \omega_n' = \omega_n - \omega_c, \end{aligned} \tag{5.2}$$

respectively. By the envelope is meant

$$E(t) = [(A \cos \omega_d t + V_c)^2 + (V_s - A \sin \omega_d t)^2]^{1/2}, \tag{5.3}$$

which is a slowly varying function of the time provided $f_d \ll f_0, f_c$ i.e., the centers of the noise and signal spectra are not too far apart in frequency. The quantities A, V_c, V_s are, of course, relatively low-frequency disturbances.

Our first task is to obtain the joint probability density of E_1 and E_2, where E_1 is the envelope at time t_0 and E_2 its value at some later time $t_0 + t$. We derive the distribution of V_{c1}, V_{c2}, V_{s1}, V_{s2}, and V_{s2} initially in rectangular coördinates. Letting

$$\begin{aligned} X_1 = V_{c1} &= \sum_{n=1}^{\infty} (a_n \cos \omega_n' t_0 + b_n \sin \omega_n' t_0); \\ X_3 = V_{c2} &= \sum_{n=1}^{\infty} [a_n \cos \omega_n' (t_0 + t) + b_n \sin \omega_n' (t_0 + t)] \\ X_2 = V_{s1} &= \sum_{n=1}^{\infty} (-a_n \sin \omega_n' t_0 + b_n \cos \omega_n' t_0); \\ X_4 = V_{s2} &= \sum_{n=1}^{\infty} [-a_n \sin \omega_n' (t_0 + t) + b_n \cos \omega_n' (t_0 + t)], \end{aligned} \tag{5.4}$$

with the help of (2.3), (2.4), and the generalized Gaussian distribution (A.3), we observe that

$$\overline{X_1^2} = \overline{X_2^2} = \overline{X_3^2} = \overline{X_4^2} = \int_0^{\infty} w(f) df \equiv \psi(0) = \mu_{11} = \mu_{22} = \mu_{33} = \mu_{44}, \tag{5.5}$$

$w(f)$ as before being the mean power spectrum of the noise; for $\overline{X_1 X_3}$ and $\overline{X_2 X_4}$ we have

$$\overline{X_1 X_3} = \overline{X_2 X_4} = \int_0^{\infty} w(f) \cos (\omega - \omega_c) t df \equiv \psi_0(t) = r_0(t)\psi = \mu_{13} = \mu_{24}; \tag{5.6}$$

(see Appendix II). It is also evident that $\overline{X_1 X_2} = \overline{X_3 X_4} = 0$, as there is no correlation between the "in-" and "out-of-phase" components of the noise. There remains finally

$$\overline{X_2 X_3} = -\overline{X_1 X_4} = \int_0^{\infty} w(f) \sin (\omega - \omega_c) t df \equiv \lambda_0(t)\psi = \mu_{23} = -\mu_{14}. \tag{5.7}$$

Accordingly, if we let $Y_1 = A \cos \omega_d t + V_c$, $Y_2 = -A \sin \omega_d t + V_s$, etc., the various Y_1, Y_2,

Y_3, Y_4 represent random variables distributed about the averages $A_0(t_0')\cos\omega_d t_0$, $-A_0(t_0')\sin\omega_d t_0$, $A_0(t_0'+t)\cos\omega_d(t_0+t)$, and $-A_0(t_0'+t)\sin\omega_d(t_0+t)$, where no correlation between the modulation and the carrier, as well as the noise, is assumed. Thus $A_1=A_0(t_0')$ represents the envelope modulation at a time t_0', not analytically related to t_0, and $A_2=A_0(t_0'+t)$ is the same modulation a time t later. Equations (5.4) become

$$\begin{aligned} X_1 &= Y_1 - A_1\cos\omega_d t_0; & X_2 &= Y_2 + A_1\sin\omega_d t_0; \\ X_3 &= Y_3 - A_2\cos\omega_d(t_0+t); & X_4 &= Y_4 + A_2\sin\omega_d(t_0+t). \end{aligned} \tag{5.8}$$

Equation (A1.3), where $s=4$, gives us the joint distribution $W_2(X_1X_2; X_3X_4; A_1A_2; t)$, for which the fundamental matrix is

$$\mathfrak{y} = \|\mu_{kl}\| = \begin{bmatrix} \psi & 0 & r_0\psi & -\lambda_0\psi \\ 0 & \psi & \lambda_0\psi & r_0\psi \\ r_0\psi & \lambda_0\psi & \psi & 0 \\ -\lambda_0\psi & r_0\psi & 0 & \psi \end{bmatrix}. \tag{5.9}$$

From (5.5)–(5.7) and from (5.9) the determinant $|\mu|$ and the cofactors μ^{kl} follow at once. Since (5.1) may be written $V(t)=Y_1\cos\omega_c t+Y_2\sin\omega_c t$, the envelopes of the wave at the two times t_0' and $t_0'+t$ are respectively

$$E_1 = (Y_1^2+Y_2^2)^{1/2} \quad\text{and}\quad E_2 = (Y_3^2+Y_4^2)^{1/2}, \tag{5.10}$$

and following the example of Rice[2] and others, we transform to polar coördinates with the help of

$$Y_1 = E_1\cos\theta_1; \quad Y_2 = E_1\sin\theta_1; \quad Y_3 = E_2\cos\theta_2; \quad Y_4 = E_2\sin\theta_2, \tag{5.11}$$

for which the Jacobian is easily shown to be E_1E_2.

The probability density W_2 becomes finally

$$\begin{aligned} &W_2(E_1E_2;\theta_1\theta_2;A_1A_2;t) \\ &= \frac{E_1E_2}{4\pi^2\psi(1-\lambda_0^2-r_0^2)}\exp\Bigg[-\frac{1}{2\psi(1-\lambda_0^2-r_0^2)}\{E_1^2+E_2^2+A_1^2+A_2^2 \\ &\quad - 2r_0E_1E_2\cos(\theta_2-\theta_1) - 2\lambda_0E_1E_2\sin(\theta_2-\theta_1) \\ &\quad - 2r_0A_1A_2\cos\omega_d t + 2\lambda_0A_1A_2\sin\omega_d t \\ &\quad + 2E_2\sin\theta_2[A_2\sin\omega_d t_2 - r_0A_1\sin\omega_d t_1] \\ &\quad + 2E_1\sin\theta_1[A_1\sin\omega_d t_1 - r_0A_2\sin\omega_d t_2] \\ &\quad + 2\lambda_0[E_2A_1\sin\theta_2\cos\omega_d t_1 - E_1A_2\sin\theta_1\cos\omega_d t_2] \\ &\quad + 2\lambda_0[E_2A_1\cos\theta_2\sin\omega_d t_1 - E_1A_2\cos\theta_1\sin\omega_d t_2] \\ &\quad - 2E_1\cos\theta_1[A_1\cos\omega_d t_1 - r_0A_2\cos\omega_d t_2] \\ &\quad - 2E_2\cos\theta_2[A_2\cos\omega_d t_2 - r_0A_1\cos\omega_d t_1]\}\Bigg], \end{aligned} \tag{5.12}$$

where $t_1=t_0$, $t_2=t_0+t$. It should be pointed out that W_2 as given by (5.12) is not

purely a probability density, in as much as averages over the phases of the modulation and over the phases of the difference frequency terms involving $\omega_d t_1$, $\omega_d t_2$ remain to be taken. Since the modulation and difference terms are not correlated, these averages are independent and, for example, may be determined in the manner of (2.12). Although periodic disturbances are not, strictly speaking, random, they may be treated as such on assuming a random distribution of the time origin with respect to an hypothetical observer. In this way the phase may be considered a random variable uniformly distributed between 0 and 2π, and a statistical average accordingly may be performed for the function containing this random variable. See section 3.10 of reference 2 for a more detailed discussion.

It is evident at once from (5.12) that the presence of a modulated signal introduces considerable mathematical complexity, and comparison with the results of Secs. 2 and 4 would indicate that this form of treatment is perhaps in most cases not so expeditious. Further, it is restricted to narrow-band disturbances, while the approach of Sections 2 to 4 is more general. However, a considerable simplification in (5.12) is possible if we observe that for $\omega' = \omega - \omega_0$, Eq. (5.7) becomes essentially zero, i.e., $\lambda_0(t) \doteq 0$, since the input spectrum is narrow-band and symmetrically distributed about $f = f_c$ and f_c heavily exceeds the effective bandwidth of the noise, as now the contribution from 0 to $-\omega_c$ nearly cancels the integral from 0 to ∞. It follows that all terms in (5.12) containing $\lambda_0(t)$ may then be safely discarded.

The case of greatest practical interest occurs when the carrier and central noise frequencies coincide; then $\omega_d = 0$, and with the usual condition $\lambda_0(t) = 0$ we have for W_2

$$W_2 = \frac{E_1 E_2}{4\pi^2 \psi^2 (1 - r^2)} \exp - (A_1^2 + A_2^2 - 2r_0 A_1 A_2)/2\psi(1 - r_0^2)$$
$$\cdot \exp - [E_1^2 + E_2^2 - 2r_0 E_1 E_2 \cos(\theta_2 - \theta_1)$$
$$- 2E_1(A_1 - r_0 A_2) \cos\theta_1 - 2E_2 \cos\theta_2 (A_2 - r_0 A_1)]/2\psi(1 - r_0^2). \qquad (5.13)$$

Notice that when $t \to \infty$ we obtain the square of the first-order density W_1. Then since $r_0(\infty) = 0$, $\lambda_0(\infty) = 0$, Eq. (5.12) transforms in the more general case where $\omega_d \neq 0$ to

$$\lim_{t \to \infty} (W_2) = W_1^2 = \left\{ \frac{E}{2\pi\psi} \exp - [A_0(t)^2 + E^2 - 2EA_0(t) \cos\phi]/2\psi \right\}^2, \qquad (5.14)$$

in which $\phi = \theta + \omega_d t$ is a new phase angle. The probability density of the envelope E is found by integrating over all phases θ, or ϕ, between 0 and 2π. The result is

$$W_1(E, A_0(t)) = (E/\psi) \exp - I_0(EA_0(t)/\psi)[A_0(t)^2 + E^2]/2\psi, \qquad (5.15)$$

a generalization of a result derived independently by Goudsmit,[27] North,[7] Rice,[2] and others, when the carrier is not modulated, i.e., when $A_0(t) = A_0$. The complete first-order probability density requires the average over the phases of the modulation, in the manner explained following (5.12). When there is no signal we have the well-known expression $W_1(E) = E\psi^{-1} \exp(-E^2/2\psi)$. Figure 4b illustrates some typical distributions of W_1 for various values of the ratio $A_0/\psi^{1/2}$, or $\sqrt{2}p^{1/2}$. As we would

[27] S. A. Goudsmit, *Comparison between signal and noise*, M.I.T. Rad. Lab. Report 43-21, Jan. 29, 1943.

expect, the presence of the signal shifts the average and the most probable values of the distribution to larger values of the envelope voltage, such that for sufficiently strong signals the most probable value of E coincides with the peak amplitude A_0.

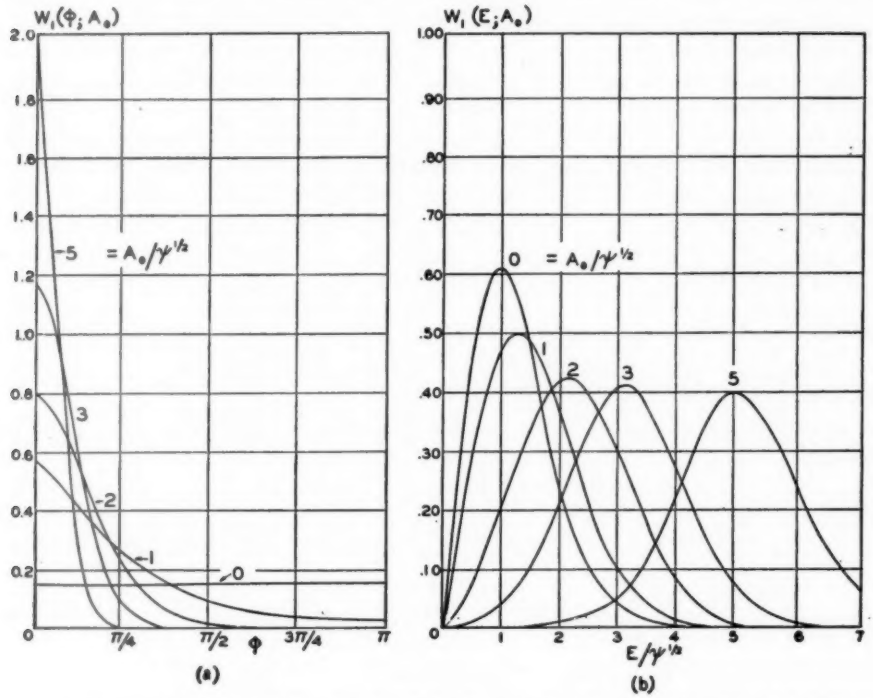


FIG. 4. Curves (a): the probability density of the phase ϕ when the carrier is unmodulated. Curves (b): the probability density of the envelope E of an unmodulated carrier and narrow-band noise, for various ratios of peak signal to r-m-s noise amplitudes, as in the case of (a).

The probability density $W_1(\theta, A_0(t))$ for the phases follows in a similar manner. From (5.14) we have after integration over E with the help of (A3.2) and (A3.7)

$$W_1(\theta, A_0(t)) = \frac{1}{2\pi} e^{-(A_0(t)\sin\phi)^2/2\psi} \left\{ \left(\frac{\pi}{2\psi}\right)^{1/2} A_0(t) \cos\phi \right.$$
$$\left. + {}_1F_1[-\tfrac{1}{2}; \tfrac{1}{2}; -(A_0(t)\cos\phi)^2/2\psi] \right\}, \qquad (5.16)$$

and again to determine the pure distribution $W_1(\theta)$ we must average over the phases of the modulation $A_0(t)$ as described above. The expression $\phi = \theta + \omega_d t$ introduced first in (5.14) is a new phase variable, shifted from the origin of θ by an amount that increases linearly with the time; Eq. (5.16) might be considered a periodically fluctuating probability density. However, the statistics of the phase θ or ϕ are not radically affected; the density functions are the same in either case, only the average values of θ and ϕ are different in the interval $0-2\pi$. Note that when there is no signal

$W_1(\theta) = 1/2\pi$, as we would expect. Figure 4a illustrates the variation of the phase density for an unmodulated carrier where the ratio $A_0/\psi^{1/2}$ takes the same values as in the example of Fig. 4b for the envelope. For no signal the distribution is uniform; as the signal strength is increased relative to the r-m-s noise voltage the phases are grouped progressively closer about the 0°-phase of the carrier, until for no noise, or what is the same thing here, for an overwhelming signal, the density becomes a delta-function at $\phi = 0$. Strong signals thus dominate the distribution of the phases for the noise, and strong signals tend to obliterate the random phasing. The curves of Fig. 4a are symmetrical about $\phi = 0$, the complete interval being $-\pi < \phi \leqq \pi$ or $0 < \phi \leqq 2\pi$.

The second-order probability density W_2 for the envelopes E_1 and E_2, or the phases θ_1 and θ_2 may be determined in the same way that W_1 was. We have finally[28]

$$W_2(E_1E_2; A_1A_2; t)$$

$$= \frac{E_1E_2}{\psi^2(1-r_0^2)} \exp - [E_1^2 + E_2^2 + A_1^2 + A_2^2 - 2A_1A_2r_0]/2\psi(1-r_0^2)$$

$$\times \sum_{m=0}^{\infty} \epsilon_m I_m\left(\frac{r_0E_1E_2}{\psi(1-r_0^2)}\right) I_m\left(\frac{A_1 - r_0A_2}{\psi(1-r_0^2)}\, E_1\right) I_m\left(\frac{A_2 - r_0A_1}{\psi(1-r_0^2)}\, E_2\right). \tag{5.17}$$

A similar but more complex result obtains from (5.12) if $\omega_d \neq 0$. When there is no modulation (5.17) may be more simply written

$$W_2(E_1E_2; t) = \frac{E_1E_2}{\psi^2(1-r_0^2)} \exp\left[\frac{-A_0^2}{\psi(1+r_0)} - (E_1^2 + E_2^2)/2\psi(1-r_0^2)\right]$$

$$\cdot \sum_{m=0}^{\infty} \epsilon_m I_m\left(\frac{r_0E_1E_2}{\psi(1-r_0^2)}\right) I_m\left(\frac{A_0E_1}{\psi(1+r_0)}\right) I_m\left(\frac{A_0E_2}{\psi(1+r_0)}\right), \tag{5.18}$$

and for no signal at all we have

$$W_2(E_1E_2; t) = \frac{E_1E_2}{\psi^2(1-r_0^2)} \exp - (E_1^2 + E_2^2)/2\psi(1-r_0^2)\, I_0\left(\frac{r_0E_1E_2}{\psi(1-r_0^2)}\right), \tag{5.19}$$

a result originally derived by Uhlenbeck[29] and used by him to determine the correlation function $R(t) = \overline{E_1E_2}$ for a half-wave linear rectifier when random noise alone is detected.

The probability density for the phases θ_1 and θ_2, $\omega_d = 0$, follows in the same way from (5.13), the final result taking the form

$$W_2(\theta_1\theta_2; A_1A_2; t)_{\omega_d=0} = [4\pi^2(1-r_0^2)]^{-1}$$

$$\exp\left[-(A_1^2 + A_2^2 - 2r_0A_1A_2)(1 - r_0^2 - \cos^2\theta_1 - \cos^2\theta_2)/2\psi(1-r_0^2)\right]$$

$$\times \sum_{m=0}^{\infty}\left[\frac{2r_0\cos(\theta_2 - \theta_1)}{1-r_0^2}\right]^m \frac{1}{m!}\left\{a_1(2\psi)^{1/2}\Gamma\left(\frac{m+3}{2}\right) {}_1F_1\left(\frac{-m}{2}; \tfrac{3}{2}; -a_1^2/4\right)\right.$$

[28] The result (5.17) has been obtained independently by Rice (communication to the author), for the less general case of an unmodulated carrier, $A_1 = A_2 = A_0$.

[29] G. E. Uhlenbeck, *Theory of the random process*, M.I.T. Laboratory Report #454, Oct. 15, 1943.

$$+\Gamma\left(\frac{m}{2}+1\right){}_1F_1\left(-\frac{m+1}{2};\tfrac{1}{2};-a_1^2/4\right)\Big\}$$

$$\times\Big\{a_2(2\psi)^{1/2}\Gamma\left(\frac{m+3}{2}\right){}_1F_1\left(-\frac{m}{2};\tfrac{3}{2};-a_2^2/4\right)$$

$$+\Gamma\left(\frac{m}{2}+1\right){}_1F_1\left(-\frac{m+1}{2};\tfrac{1}{2};-a_2^2/4\right)\Big\}, \tag{5.20}$$

$a_1=b_1\cos\theta_1$, $a_2=b_2\cos\theta_2$, (for b_1, b_2 see Sec. 6).

6. The correlation function and mean power for the envelope of a signal and noise. The results of the preceding section may be used in (1.7b) to give us an expression for the correlation function and the mean power associated with the *envelope* of an incoming signal and noise wave. Our primary interest is with the low-frequency output, or detected envelope, but the general treatment will first be outlined in brief fashion below, before returning to a more detailed examination of the former.

For the general non linear device we follow the suggestion of Rice (ref. 2, Sec. 4.3) and use the Fourier transform (1.8) to represent our output current $I=g(V)$ as a function of the input disturbance $V=E\cos(\omega_c t-\theta)$. Expanding the exponential in a series of Bessel functions gives us

$$I=\sum_{l=0}^{\infty}I_l=\sum_{l=0}^{\infty}B_l(E)\cos l(\omega_c t-\theta), \tag{6.1}$$

where $B_l(E)$ represents the envelope of the lth output band, viz:

$$B_l(E)=\frac{i^l\epsilon_l}{2\pi}\int_C f(iz)J_l(Ez)dz, \tag{6.2}$$

which in principle may be observed when all other contributions are eliminated by an appropriate band-pass filter, centered about the lth zone and followed by a linear half-wave rectifier or "envelope tracer," as it is sometimes called. If $W_1(E)$ is the probability density of the input envelope, the density function for the lth region is

$$D_{1l}(E)=B_l(E)W_1(E)=W_1(B_l)B_l(E)dB_l/dE, \tag{6.3}$$

since

$$W_1(B_l)=W_1(E)dE/dB_l, \tag{6.4}$$

where $W_1(B_l)$ is the density function of B_l. This latter quantity, $W_1(B_l)$, may be found from (5.15) and from (6.2) by differentiation. For higher order densities we have

$$W_2(B_{1l},B_{2l};t)=W_2(E_1E_2;t)\Big/\left|\frac{\partial(B_{1l}B_{2l})}{\partial(E_1E_2)}\right|,\ \text{etc.} \tag{6.4a}$$

An important consequence of (6.4) and (6.4a) is that these results enable us to determine the probability density of the envelope of a distribution after successive non linear operations, *provided*, of course, that in each operation the concept of the envelope remains, i.e., we have a narrow band wave undergoing rectification and not one whose mean frequency is comparable with the carrier.

The correlation function and the mean power may be written in a similar manner. We have from (2.12), (5.12), and (6.1) for the complete correlation

$$R(t) = \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} T_0'^{-1}\int_0^{T_0'} dt_0' \int_0^{T_0} T_0^{-1} dt_0 \int_0^{\infty} dE_1 \int_0^{\infty} dE_2 \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 B_l(E_1) B_m(E_2)$$
$$W_2(E_1E_2;\, \theta_1\theta_2;\, A_1A_2;\, t) \cos l(\omega_c t_0 - \theta_1) \cos m[\omega_c(t_0 + t) - \theta_2], \tag{6.5}$$

which simplifies considerably when the noise band is symmetrical, i.e., $\lambda_0 \doteq 0$, the carrier is unmodulated, and $\omega_d = 0$. The spectrum may be found from (1.6). The mean power associated with the lth harmonic region is given when $t=0$ in (6.5), but an alternative and sometimes simpler expression may be obtained with the help of $W_1(E)$ and (2.33a), applied to (6.1), (6.2) after averaging over the phases of the zone central frequencies $l\omega_c$.

Specifically, let us examine the interesting case of the νth-law half-wave rectifier when a modulated signal and narrow-band noise together enter the apparatus. This is essentially the receiver problem, as we are concerned only with the low-frequency output—namely, the modified envelope. Then it is in keeping with the problem to set $\omega_d = \lambda_0(t) = 0$, the latter on the assumption of a symmetrical spectrum, and we find for the noise contribution to the correlation function as indicated by the subscript N, *cf.* 2(.11), from (5.17) in (6.5),

$$R(t)_N = \int_0^{\infty} dE_1 \int_0^{\infty} dE_2 B_0(E_1) B_0(E_2) W_2(E_1E_2;\, A_1A_2;\, t)$$
$$= \gamma^2 f(A_1A_2;\, r_0) \sum_{m=0}^{\infty} \epsilon_m \int_0^{\infty} dE_1 \int_0^{\infty} dE_2 (E_1E_2)^{\nu+1} I_m\left(\frac{E_1E_2r_0}{\psi(1-r_0^2)}\right) I_m(b_1E_1) I_m(b_2E_2)$$
$$\times e^{-(E_1^2+E_1^2)/2\psi(1-r_0^2)} \tag{6.6}$$

where $f(A_1, A_2; r_0)$ is given by

$$f(A_1, A_2; r_0) = \psi^2(1-r_0^2)]^{-1} \exp\,[(-A_1^2 - A_2^2 + 2r_0A_1A_2)/2\psi(1-r_0^2)], \quad \text{and}$$

$$b_1 = \frac{A_1 - r_0A_2}{\psi(1-r_0^2)}; \qquad b_2 = \frac{A_2 - r_0A_1}{\psi(1-r_0^2)}; \qquad B_0(E_1) = \gamma E_1^{\nu}, \text{ etc.} \tag{6.6a}$$

The scale factor γ is related to the tube factor β used throughout this paper by

$$\gamma = \frac{\beta}{2\pi}\int_0^{\pi} \sin^{\nu}\theta d\theta = \frac{\beta}{2\sqrt{\pi}} \frac{\Gamma[(\nu+1)/2]}{\Gamma[(\nu/2+1)]}. \tag{6.7}$$

This follows because in (6.6) we are dealing with the *envelope* E rather than the amplitude V of the wave. Now the output current I consists of the positive halves of the oscillations of βV^{ν}, while the envelope of I is the same as that of βV^{ν}. But the area under the loops of I is to the area under βE^{ν} as the area under a loop of $\sin^{\nu}\theta$ is to an area of unit height and length 2π, so that as far as the low-frequency portion of the output is concerned the loops of I are "smeared" together into a current which varies as γE^{ν}, with γ given by (6.7) above. Compare the appropriate results of the present section with those of the next, where the problem is treated from the point of view of the instantaneous amplitude.

There appear to be three principal ways of evaluating $R(t)_N$, each yielding results different in form and usefulness. The first is achieved with the aid of the transformation

$$E_1 = [\psi(1-r_0^2)]^{1/2} z^{1/2} e^{\phi/2}, \qquad E_2 = [\psi(1-r_0^2)]^{1/2} z^{1/2} e^{-\phi/2}, \tag{6.8}$$

followed by the successive application of the expression for the product of two Bessel functions (*cf.* (2), p. 148, ref. 21), the integral form of the modified Bessel function of the second kind ((7), p. 182, ref. 21), of argument $n-2k$, and finally, by ((11), p. 410, ref. 21). The second form follows after expanding $I_m(E_1E_2r_0/\psi(1-r_0^2))$ and using termwise integration, with the help of (A3.7), and the third employs a contour integral representation of the Bessel functions in (6.6) applied to the addition formula for $I_0[(x^2+y^2+2xy\cos\phi)^{1/2}]$, along with a reversal of the order of integration. We give the final results:

For the first method we have

$$\begin{aligned} R(t)_N = {} & \gamma^2 2^\nu \psi^\nu (1-r_0^2)^{\nu+1} e^{-[A_1^2+A_2^2-2r_0A_1A_2]/2\psi(1-r_0^2)} \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\epsilon_m r_0^m c_1^{m+2n} c_2^m}{2^{m+\nu} m!} \\ & \times \sum_{k=0}^{n} \frac{(c_1^{-1}c_2)^{2k}\Gamma(\nu/2+k+m+1)\Gamma(\nu/2-k+m+n+1)}{k!(k+m)!(n-k)!(m+n-k)!} \\ & \times {}_2F_1\left(\frac{\nu}{2}+k+m+1;\ \frac{\nu}{2}-k+m+n+1;\ m+1;\ r_0^2\right), \end{aligned} \tag{6.9}$$

$$c_1 = b_1[\psi(1-\nu_0^2)]^{1/2},\ c_2 = b_2[\psi(1-\nu_0^2)]^{1/2}.$$

In the case of modulation we must now apply (2.12) to determine the average over the phases $\omega_A t_0'$. As one can readily see from (6.9) the effort is formidable, even for the simplest modulations. However, a less general but nonetheless important case arises when $A_1=A_2=A_0$, a constant. With the assistance of (A3.20b) we obtain for the complete low-frequency correlation function

$$\begin{aligned} R(t) = {} & \gamma^2 2^\nu \psi^\nu e^{-A_0^2/\psi(1+r_0)} \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\epsilon_m r_0^m p^{m+n}}{(1+r_0)^{2m+2n}} \\ & \times \sum_{k=0}^{n} \frac{\Gamma(\nu/2+k+m+1)\Gamma(\nu/2-k+m+n+1)}{k!(k+m)!(n-k)!(m+n-k)!} \\ & \times {}_2F_1\left(k-\frac{\nu}{2}-n,\ -k-\frac{\nu}{2};\ m+1;\ r_0^2\right), \qquad p \equiv A_0^2/2\psi. \end{aligned} \tag{6.10}$$

For noise alone the low-frequency correlation is at once found from (6.10), since only the terms for which $m=n=0$ contribute:

$$\begin{aligned} R(t)_{\text{noise}} &= \gamma^2 2^\nu \psi^\nu \Gamma\left(\frac{\nu}{2}+1\right)^2 {}_2F_1\left(-\frac{\nu}{2},\ -\frac{\nu}{2};\ 1;\ r_0^2\right) \\ &= \frac{\psi^\nu \beta^2 2^{\nu-2}}{\pi}\,\Gamma\left(\frac{\nu+1}{2}\right)^2 {}_2F_1\left(-\frac{\nu}{2},\ -\frac{\nu}{2};\ 1;\ r_0^2\right), \end{aligned} \tag{6.11}$$

the latter from (6.7). For the second method we write

$$R(t)_N = \gamma^2 2^\nu (1-r_0^2)^{\nu+1}\psi^\nu \exp\{-(A_1^2+A_2^2-2r_0A_1A_2)/2\psi(1-r_0^2)\}$$
$$\times \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\epsilon_m r_0^{2n+m}\Gamma(m+n+\nu/2+1)^2(c_1c_2)^m}{(m!)^2 n!(n+m)!2^m}$$
$$\times {}_1F_1\left(\frac{\nu}{2}+m+n+1;\, m+1;\, \frac{c_1^2}{2}\right){}_1F_1\left(\frac{\nu}{2}+m+n+1;\, m+1;\, \frac{c_2^2}{2}\right), \quad (6.12)$$

equivalent to (6.9), while the counterpart of (6.10) is

$$R(t) = \gamma^2 2^\nu \psi^\nu (1-r_0^2)^{\nu+1}e^{-p}\sum_{m=0}^{\infty}\frac{\epsilon_m r_0^m}{(m!)^2}\left(\frac{1-r_0}{1+r_0}\right)^m p^m \sum_{n=0}^{\infty}\frac{r_0^{2n}\Gamma(m+n+1+\nu/2)^2}{n!(n+m)!}$$
$$\times {}_1F_1\left[-n-\frac{\nu}{2};\, m+1;\, -p\frac{(1-r_0)}{1+r_0}\right]^2 \quad (6.13)$$

from (A3.2). When there is no signal it is easy to verify that (6.13) reduces to (6.11), again with the help of (A3.20b).

There remains the third approach. This enables us to write finally, for the correlation function

$$R(t)_N = \gamma^2 2^\nu \psi^{\nu+2}(1-r_0^2)^{\nu+2} f(A_1A_2;\, r_0)\Gamma(\nu/2+1)^2 \sum_{m=0}^{\infty}\frac{(\nu/2+1)_m^2}{(m!)^2}$$
$$\times \frac{1}{2\pi i}\int_{\mathbf{D}} \frac{\xi^{\nu+1}e^{\xi}[\xi r_0 + c_1c_2/2]^{2m}d\xi}{[\xi - c_1^2/2]^{\nu/2+1}[\xi - c_2^2/2]^{\nu/2+1}}. \quad (6.14)$$

The path **D** of integration in the ξ-plane is taken to be sufficiently large so that the series in the integrand converge satisfactorily. The contour also includes any branch points and poles of the integrand of (6.14). Expanding the denominator of (6.14) in series and using the binomial development of $(\xi r_0 + c_1c_2/2)^{2m}$, gives, after termwise integration for the case when there is no modulation, viz., $A_1 = A_2 = A_0$, and $c_1 = c_2 = c$, $c^2/2 = p(1-r_0)/(1+r_0)$,

$$R(t) = 2^\nu \psi^\nu \gamma^2 \Gamma\left(\frac{\nu}{2}+1\right)^2 (1-r_0^2)^{\nu+1}e^{-p}\sum_{m=0}^{\infty}\frac{(\nu/2+)_m^2(2m)!}{m!^2}$$
$$\times \sum_{n=0}^{2m}\frac{r_0^{2m-n}p^n(1-r_0)^n}{n!^2(2m-n)!(1+r_0)^n}$$
$$\times {}_1F_1[-2m-\nu-1+n;\, n+1;\, -p(1-r_0)/(1+r_0)]. \quad (6.15)$$

A somewhat different form of the same expression is found by making the substitution $z = \xi - c^2/2$, and again expanding $(\xi r_0 + c^2/2)^{2m}$ in a binomial series, followed by termwise integration of the resulting series. The low-frequency correlation is then

$$R(t) = 2^\nu \psi^\nu \gamma^2 \Gamma\left(\frac{\nu}{2}+1\right)(1-r_0^2)^{\nu+1}e^{-p}\sum_{m=0}^{\infty}\frac{(\nu/2+1)_m^2(2m)!}{m!^2}\sum_{n=0}^{2m}\frac{(1-r_0)^n r_0^{2m-n}p^n}{(2m-n)!n!}$$
$$\times \sum_{k=0}^{\infty}\frac{1}{k!}\left[\frac{p(1-r_0)}{1+r_0}\right]^{1+\nu-k}\frac{\Gamma(\nu+2)}{\Gamma(\nu-k+2)\Gamma(\nu+n-k+2)}. \quad (6.16)$$

Notice that when ν is integral, the series in k terminates after $\nu+2$ terms.

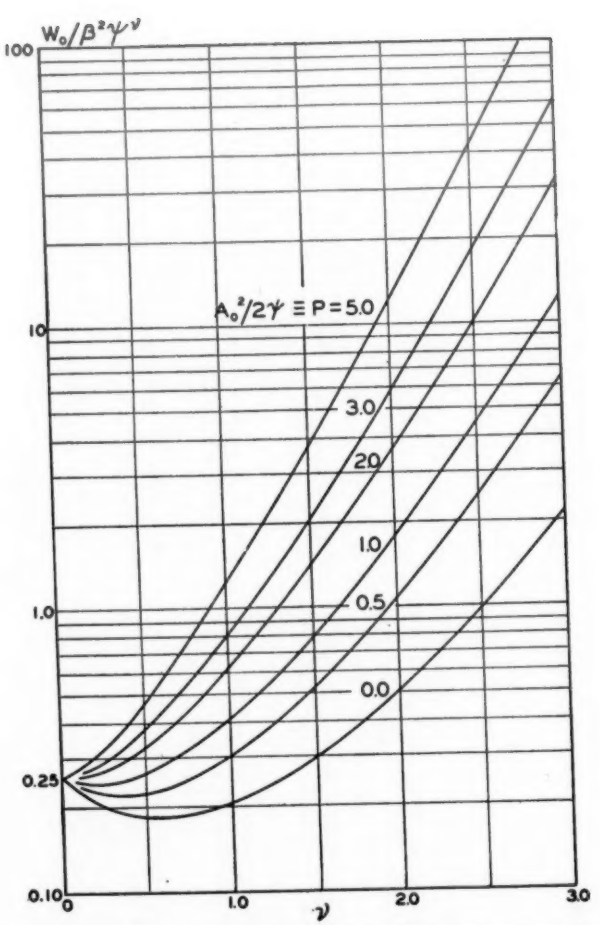


FIG. 5. Mean power output in the low-frequency components following a half-wave νth-law rectifier, for a number of values of the parameter $p = A_0^2/2\psi$.

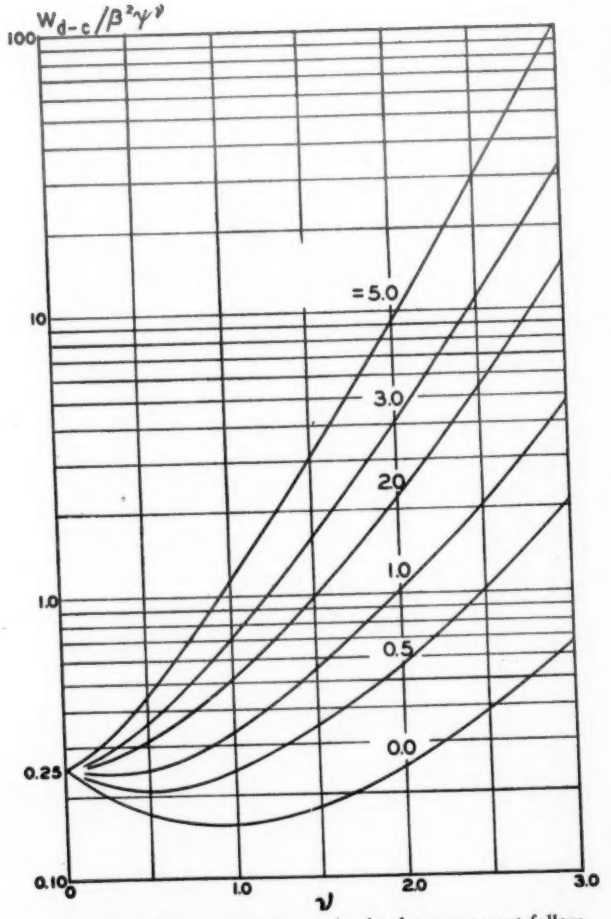


FIG. 6. Mean power output in the d-c component following a half-wave νth-law rectifier for various values of the parameter $p = A_0^2/2\psi$.

The mean total low-frequency power $(W_\tau)_0$ and the power in the $d-c$ component W_{d-c} may be obtained respectively for the various cases considered above on setting $t=0$ and $t=\infty$ in $R(t)$. However, a simpler way of obtaining the same results is to use (6.6) or (6.7) with (5.15) directly. Application of (A3.7) then gives us

$$(W_\tau)_0 = \overline{I_0^2} = \gamma^2 2^\nu \psi^\nu \Gamma(\nu+1)\,{}_1F_1(-\nu;1;-p), \qquad A_1 = A_2 = A_0, \tag{6.17}$$

and for the $d-c$

$$W_{d-c} = (\bar{I}_0)^2 = \gamma^2 2^\nu \psi^\nu \Gamma\left(\frac{\nu}{2}+1\right)^2 {}_1F_1\left(-\frac{\nu}{2};1;-p\right)^2. \tag{6.18}$$

When the carrier is modulated we have only to average the various $p(i_0')^j$ that are present in the above. Curves illustrating $(W_\tau)_0$ and W_{d-c} as a function of the law of the detector are included in Figs. 5 and 6, for a number of values of $p=A_0^2/2\psi$. It is interesting to note that the output powers actually decrease as ν increases in the interval $0 \backsim \nu \backsim 1.0$ when the signal is weak relative to the noise. Strong signals suppress the noise, and the output is then proportional to p^ν for both total-, low-frequency, and $d-c$ powers, namely,

$$(W_\tau)_0 \cong \gamma^2 2^\nu \psi^\nu p^\nu (1+\nu^2/p+\cdots) \text{ and } W_{d-c} \cong \gamma^2 2^\nu \psi^\nu p^\nu (1+\nu^2/2p+\cdots), \tag{6.19}$$

from (A3.3).

7. General half-wave rectification of random noise. A relatively simple case of considerable interest arises when random noise alone enters a non linear device which passes only the positive amplitudes of the incoming disturbance. The current-voltage characteristic assumes the following general form:

$$\begin{aligned} I &= \beta V^\nu, \qquad V>0, \qquad \nu>0 \\ &= 0, \qquad V<0. \end{aligned} \tag{7.1}$$

where we restrict ourselves to unsaturated cases, corresponding in practice to (relatively) small-signal rectification. Here V represents the instantaneous input noise voltage, and I is the output current.

The analysis of the present section is more general than that of Sections 5 and 6 in that it is capable of handling broad-band noise, where the concept of the envelope is no longer meaningful; it is less general in that signals are excluded from the discussion. Also, a more detailed study of the higher order spectral regions is included, for the case of narrow-band noise. In fact, Section 7 is in itself a study of a special case of the analysis of Sections 3 and 4, when there is no bias and no carrier.

(*a*) *Broad-band Noise.* We consider first the case of wide-band noise, whose central frequency, spectrally speaking, is comparable to the bandwidth of the disturbance. See Fig. 7(a) or (b). From (2.9), (2.11), and (7.1) the correlation function of the output in the absence of a signal is

$$\begin{aligned} R(t) = \frac{\beta^2\psi^\nu}{2\pi(1-r^2)^{1/2}} \int_0^\infty dX_1 \int_0^\infty dX_2 (X_1X_2)^\nu \\ \times \exp\left[-(X_1^2+X_2^2-2rX_1X_2)/2\psi(1-r^2)\right] \end{aligned} \tag{7.2}$$

where $r(t)$ is the normalized input correlation, cf. Eq. (A2.7). The integration of (7.2) may be effected directly by using polar coördinates: $x=\rho\cos\theta$, $y=\rho\sin\theta$. The latter approach gives us for (7.2)

$$R(t) = \frac{\beta^2\psi^\nu 2^{-\nu-1}}{[2\pi(1-r^2)^{1/2}]}\int_0^\pi \sin^\nu\phi d\phi\int_0^\infty \rho^{2\nu+1}e^{-a\rho^2}d\rho,$$

$$a = (1-r\sin\phi)/2(1-r^2)$$

$$= \frac{\beta^2\psi^\nu\Gamma(\nu+1)(1-r^2)^{\nu+1/2}}{4\pi}\int_0^\pi \frac{\sin^\nu\phi d\phi}{(1-r\sin\phi)^{\nu+1}}, \tag{7.3}$$

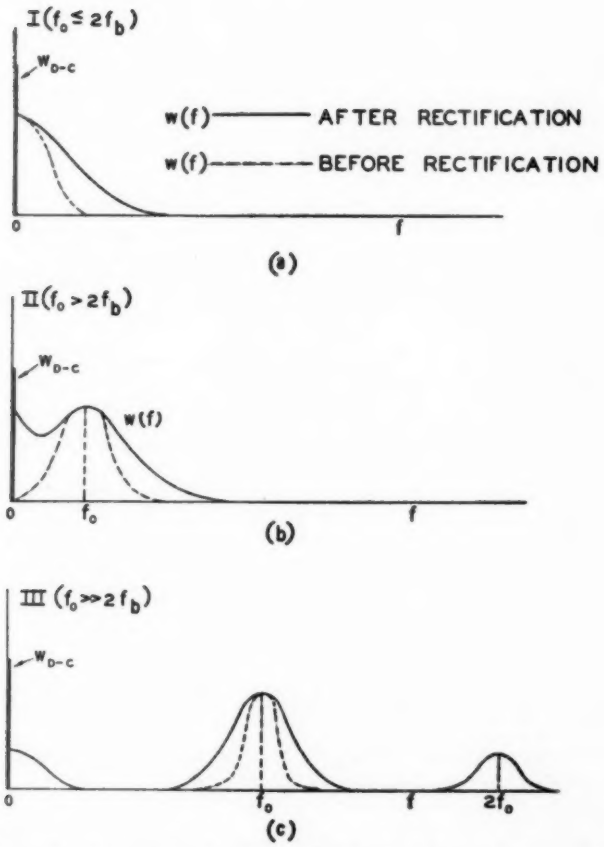


FIG. 7. Examples of narrow-band and broad-band noise before and after rectification. The normalization is arbitrary.

where the substitution $2\theta=\phi$ has been made. Expansion of the denominator and term-wise integration (since $r\sin\phi\leqq 1$), with the help of

$$\sin^{\nu}\phi(1 - r\sin\phi)^{-(\nu+1)} = \sum_{n=0}^{\infty} \frac{r^n(\sin\phi)^{\nu+n}\Gamma(\nu + n + 1)}{n!\Gamma(\nu + 1)} \quad \text{and}$$

$$\int_0^{\pi} \sin^{\nu+n}\phi d\phi = \frac{\sqrt{\pi}\,\Gamma[(\nu + n + 1)/2]}{\Gamma[(\nu + n)/2 + 1]} \tag{7.4}$$

yields

$$\begin{aligned} R(t) &= \frac{\beta^2\psi^{\nu}(1 - r^2)^{\nu+1/2}}{4\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{r^n\Gamma[(\nu + n + 1)/2]\Gamma(\nu + n + 1)}{n!\Gamma[(\nu + n)/2 + 1]} \\ &= \frac{\beta^2 2^{\nu-2}\psi^{\nu}}{\pi}(1 - r^2)^{\nu+1/2}\left\{\Gamma\left(\frac{\nu + 1}{2}\right)^2 {}_2F_1\left(\frac{\nu + 1}{2}, \frac{\nu + 1}{2}; \tfrac{1}{2}; r^2\right)\right. \\ &\quad \left. + 2r\Gamma\left(\frac{\nu}{2} + 1\right)^2 {}_2F_1\left(\frac{\nu}{2} + 1, \frac{\nu}{2} + 1; \tfrac{3}{2}; r^2\right)\right\} \\ &= \frac{\beta^2 2^{\nu}\psi^{\nu}}{4\pi}\left\{\Gamma\left(\frac{\nu + 1}{2}\right)^2 {}_2F_1\left(-\frac{\nu}{2}, -\frac{\nu}{2}; \tfrac{1}{2}; r^2\right)\right. \\ &\quad \left. + 2r\Gamma\left(\frac{\nu}{2} + 1\right)^2 {}_2F_1\left(\frac{1 - \nu}{2}, \frac{1 - \nu}{2}; \tfrac{3}{2}; r^2\right)\right\}, \end{aligned} \tag{7.5}$$

this last with the help of (A3.20b). The series form of (7.5) is more convenient when the spectrum is desired; $R(t)$ may be written accordingly

$$\begin{aligned} R(t) = \frac{\beta^2 2^{\nu}\psi^{\nu}}{4\pi} \sum_{n=0}^{\infty} \Bigg[& \frac{(-\nu/2)_n^2\Gamma[(\nu + 2)/2]^2 2^{2n} r^{2n}}{(2n)!} \\ & + \frac{[(1 - \nu)/2]_n^2\Gamma[\nu/2 + 1]^2 2^{2n+1} r^{2n+1}}{(2n + 1)!}\Bigg]. \end{aligned} \tag{7.6}$$

A number of interesting special cases follow from (7.6). For half-wave linear rectifiers ($\nu = 1$) we obtain

$$\begin{aligned} R(t)_{\nu=1} &= \frac{\beta^2\psi}{2\pi}\left({}_2F_1(-\tfrac{1}{2}, -\tfrac{1}{2}; \tfrac{1}{2}; r^2) + \frac{\pi r}{2}\right) \\ &= \frac{\beta^2\psi}{2\pi}\left\{r\left(\sin^{-1} r + \frac{\pi}{2}\right) + (1 - r^2)^{1/2}\right\}, \end{aligned} \tag{7.7}$$

a result derived independently by Van Vleck, North, and Rice (see Sec. 4.7 of reference 2). Half-wave quadratic detectors ($\nu = 2$) give us

$$\begin{aligned} R(t)_{\nu=2} &= \frac{\beta^2\psi^2}{2\pi}\left(\frac{\pi}{2}(1 + 2r^2) + 4r\,{}_2F_1(-\tfrac{1}{2}, -\tfrac{1}{2}; \tfrac{3}{2}; r^2)\right) \\ &= \frac{\beta^2\psi^2}{2\pi}\left\{\left(\frac{\pi}{2} + \sin^{-1} r\right)(1 + 2r^2) + 3r(1 - r^2)^{1/2}\right\}. \end{aligned} \tag{7.8}$$

We may continue in this fashion when ν is integral and write $R(t) = f(r, (1 - r^2)^{1/2},$

$\sin^{-1}r$) with the help of the recurrence relations for the hypergeometric function.

The total mean output power W_τ, the d-c power W_{d-c}, and the mean total a-c power W_{a-c} all follow at once from (2.26d), (2.33), (7.5), and (A3.20a). We have finally

$$W_\tau = \frac{\beta^2\psi^\nu 2^{\nu-1}}{\sqrt{\pi}}\Gamma(\nu+\tfrac{1}{2}); \quad W_{d-c} = \frac{\beta^2\psi^\nu 2^{\nu-2}}{\pi}\Gamma\left(\frac{\nu+1}{2}\right)^2; \quad W_{a-c} = W_\tau - W_{d-c} \tag{7.9}$$

Curves of W_τ, W_{d-c}, and W_{a-c} are shown in Fig. 8. It is evident that characteristics for which ν is large (>2) exhibit outputs which are chiefly a-c. The spectra associated with the output may be found from (1.6) and (7.6), where integrals of the type

$$c_{0,n}(f) = 4\int_0^\infty r(t)^n \cos \omega t dt, \qquad \omega = 2\pi f,$$

must be considered. Observe that when ν is an even integer, the first series in (7.6) terminates after the term for which $n=\nu/2$, and when ν is odd, after $n=(\nu-1)/2$. In particular, the mean output spectrum for the linear rectifier is

$$W(f)_{\nu=1} = \frac{\beta^2\psi}{2\pi}\left\{2\delta(f-0) + \frac{\pi}{2}c_{0,1}(f) + \sum_{n=1}^{\infty}\frac{(2n)!c_{0,2n}(f)}{n!^2 2^{2n}(2n-1)^2}\right\}, \tag{7.10}$$

and for a quadratic detector

$$W(f)_{\nu=2} = \frac{\beta^2\psi^2}{2\pi}\left\{\pi\delta(f-0) + 4c_{0,1}(f) + \pi c_{0,2}(f) + 4\sum_{n=1}^{\infty}\frac{(2n)!c_{0,2n+1}(f)}{[n!2^{2n}(2n-1)]^2(2n+1)}\right\}, \tag{7.11}$$

where $\delta(f-0)$ is the familiar delta-function. Curves illustrating (7.10) and (7.11) when the input spectrum is Gaussian, Eq. (A2.9), $\omega_c=0$, are given in reference 9.

(*b*) *Narrow-band Noise.* The output now consists, as explained above in more detail in section 3, of bands of noise located about harmonics of the central frequency f_c. The spectra associated with these harmonic zones are "smeared out," or distorted from their original shape, as Fig. 7(c) indicates. Our result (7.6) is expressed in a form convenient for obtaining the correlation functions associated with the various harmonic regions lf_c, $l=0, 1, 2, 3, \cdots$, and hence the spectral distribution of the mean power and the mean total power itself pertaining to these regions.

Here the input correlation function is given by (A2.8). When this is substituted into (7.6) we obtain series involving $(\cos \omega_c t)^{2n}$ and $(\cos \omega_c t)^{2n+1}$. The expansion in multiples of $\omega_c t$ is then used, and the correlation function of the lth zone follows from (7.6) on taking those values of n such that $j=(2n-l)/2(\geqq 0)$ is integral for even l, and $j=(2n-l+1)/2(\geqq 0)$ for odd values of $l(>0)$. We may write finally

$$R_l(t) = \pi^{-1}\beta^2\psi^\nu 2^{\nu-2}\Gamma\left(\frac{\nu+1}{2}\right)^2\epsilon_l \cos l\omega_c t \sum_{n=l/2}^{\infty}\frac{(-.\nu/2)_n^2 r_0(t)^{2n}}{(n-l/2)!(n+l/2)!}, \qquad l = 0, 2, 4, \cdots, \tag{7.12a}$$

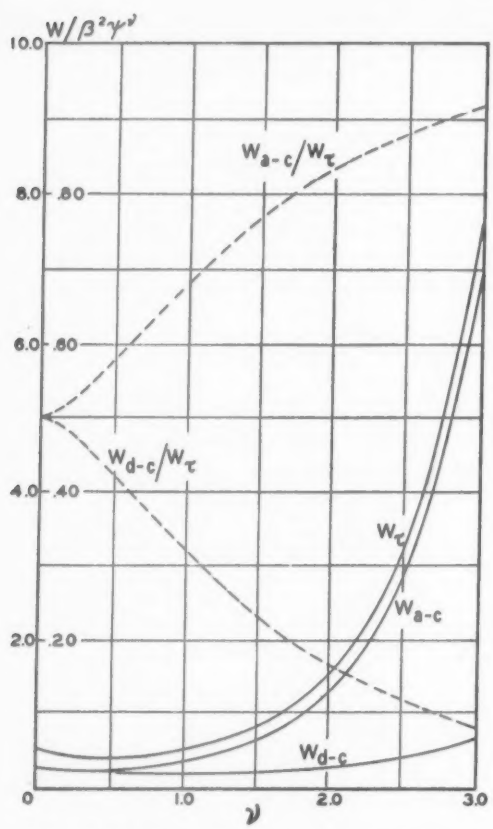


FIG. 8. The mean total, a-c, and d-c power outputs of a half-wave νth-law detector when the incoming disturbance is broadband noise, as a function of ν. The right-hand scale applies for the ratios.

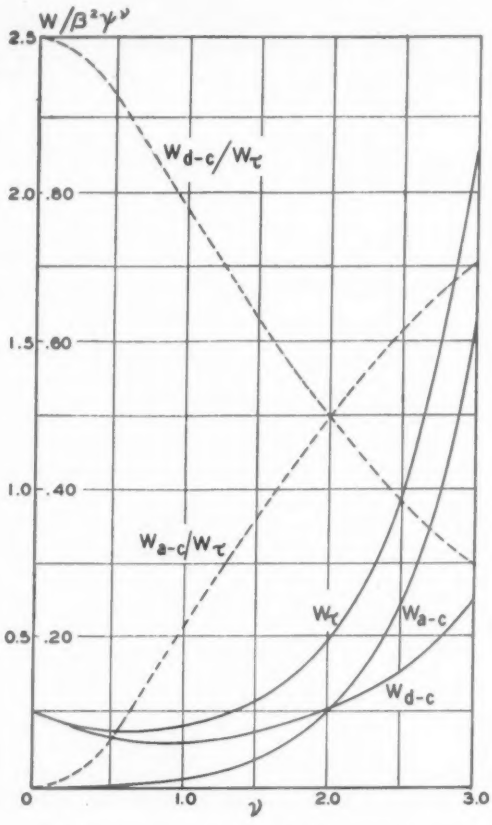


FIG. 9. The mean total and a-c powers associated with the low-frequency output when the incoming random noise is narrow-band, as a function of ν. The d-c power output is the same as that in Fig. 8, and the right-hand scale applies in the case of the ratios.

and

$$R_l(l) = \pi^{-1}\beta^2\psi^\nu 2^{\nu-1}\Gamma\left(\frac{\nu}{2}+1\right)^2 \cos l\omega_c l \sum_{n=(l-1)/2}^{\infty} \frac{[(1-\nu)/2]_n^2 r_0(l)^{2n+1}}{[(2n-l+1)/2]![(2n+l+1)/2]},$$

$$l = 1, 3, 5, \cdots . \qquad (7.12b)$$

We observe that contributions to the spectrum about the frequency lf_c arise from terms for which $n \geqq l/2, (l-1)/2$; values of n less than this do not appear. Now (7.12a) and (7.12b) may be summed; for even values of l we set $n-l/2=m$, $m=0, 1, 2, \cdots$, and find that

$$\sum_{n=l/2}^{\infty} \frac{[-\nu/2]_n^2 r_0^{2n}}{(n-l/2)!(n+l/2)!} = r_0^l \sum_{m=0}^{\infty} \frac{[-\nu/2]_{m+l/2}^2 r_0^{2m}}{m!(m+l)!}$$

$$= \frac{r_0^l(-\nu/2)_{l/2}^2}{l!} \sum_{m=0}^{\infty} \frac{[(l-\nu)/2]_m^2 r_0^{2m}}{m!(l+1)_m}$$

$$= \frac{r_0^l[-\nu/2]_{l/2}^2}{l!} {}_2F_1\left(\frac{l-\nu}{2}, \frac{l-\nu}{2}; l+1; r_0^2\right), \quad (7.13)$$

and the correlation function appropriate to the even harmonics is then

$$R(l) = \frac{\beta^2\psi^\nu 2^{\nu-2}}{\pi} \epsilon_l \left(-\frac{\nu}{2}\right)_{l/2}^2 \frac{\Gamma[(\nu+1)/2]^2}{l!} r_0^l \cos l\omega_c t$$

$$\times {}_2F_1\left(\frac{l-\nu}{2}, \frac{l-\nu}{2}; l+1; r_0^2\right), \qquad l = 0, 2, 4, \cdots. \quad (7.14)$$

A similar procedure for odd values of l, letting $n-(l-1)/2=m$, $m=0, 1, 2 \cdots$ in (7.12b) yields

$$R_l(l) = \frac{\beta^2\psi^\nu 2^{\nu-1}}{\pi} \left(\frac{1-\nu}{2}\right)_{(l-1)/2}^2 \frac{\Gamma[\nu/2+1]^2}{l!} r_0^l \cos l\omega_c t$$

$$\times {}_2F_1\left(\frac{l-\nu}{2}, \frac{l-\nu}{2}; l+1; r_0^2\right), \qquad l = 1, 3, 5, \cdots. \quad (7.15)$$

Notice that when the detector characteristic is an even power of the incoming wave, only those zones for which l is even and equal to or less than ν (even) are produced, while all *odd* harmonic zones ($l=1, 3, 5, 7 \cdots$) appear in the output. A like situation is encountered in odd-powered characteristics: only the zones for which $0<l\leqq\nu$ (l and ν odd) are generated, whereas all even regions exist. The above is analogous to the half-wave rectification of a sine wave if we identify the narrow-band input noise with a sinusoid having the frequency $f_0=f_c$. The amplitudes, and hence the powers, of the zones about lf_c corresponding to the sinusoidal components $lf_0=lf_c$ are *not* equal, as a Fourier analysis readily shows, but their number and location in the spectrum are.

Of particular interest in practice is the low-frequency output $l=0$. The complete low-frequency correlation function is seen from (7.14) to be just (6.11), which in the instance $\nu=1$ may also be expressed in terms of the complete elliptic integrals E and K of modulus r_0:

$$R_0(t)_{\nu=1} = \frac{\beta^2\psi}{\pi^2}\left\{E(r_0) - \tfrac{1}{2}(1-r_0^2)K(r_0)\right\}. \quad (7.16)$$

For the quadratic rectifier the form of $R_0(t)_{\nu=2}$ is quite simple:

$$R_0(t)_{\nu=2} = \beta^2\psi^2(1+r_0^2)/4. \quad (7.17)$$

When ν is even we obtain a polynomial in r_0, viz:

$$R_0(t)_{\nu=2n} = \beta^2\psi^{2n}\left[\frac{(2n)!}{2^{n+1}n!}\right]^2 {}_2F_1(-n, -n; 1; r_0^2)$$

$$= \left[\frac{\beta\psi^n(2n)!}{2^{n+1}n!}\right]^2 \sum_{j=0}^{n} {}_nC_j^2\, r_0^{2(n-j)}. \tag{7.18}$$

The d-c component may be found when j is set equal to n.

Again we may determine the various powers by setting $t=0$ or $t=\infty$ in (7.14) and (7.15) and using (A3.20a). The mean total power in the low-frequency region and in the d-c are respectively

$$(W_\tau)_0 = R_0(0) = \frac{\beta^2\psi^\nu 2^{\nu-2}}{\pi}\,\frac{\Gamma(\nu+1)\Gamma[(\nu+1)/2]^2}{\Gamma(\nu/2+1)^2};$$

$$W_{d-c} = R_0(\infty) = \frac{\beta^2\psi^\nu 2^{\nu-2}}{\pi}\,\Gamma\left(\frac{\nu+1}{2}\right)^2, \tag{7.19}$$

the latter agreeing with our previous result (7.9), as we would expect in virtue of the general theorem stated at the end of Appendix II. The higher ($l>0$) spectral regions are pure a-c. The mean a-c power as in general

$$(W_{a-c})_l = R_l(0 = \frac{\beta^2\psi^\nu 2^{\nu-2}}{\pi}\epsilon_l\left(-\frac{\nu}{2}\right)_{l/2}^2 \frac{\Gamma[(\nu+1)/2]^2\Gamma(\nu+1)}{\Gamma[(\nu+l)/2+1]^2},$$

$$l = 2, 4, \cdots \tag{7.20a}$$

and

$$(W_{a-c})_l = \frac{\beta^2\psi^\nu 2^{\nu-1}}{\pi}\left(\frac{1-\nu}{2}\right)_{(l-1)/2}^2 \frac{\Gamma(\nu/2+1)^2\Gamma(\nu+1)}{\Gamma[(\nu+l)/2+1]^2}, \quad l = 1, 3, 5, \cdots. \tag{7.20b}$$

The contributions from all harmonic zones to the total output power $\sum_l R_l(0)$ is simply the first equation in (7.9). Figures 9 and 10 show W_l as a function of ν for several different values of l. It is interesting to note from Fig. 9 that when $\nu<2$ the d-c power exceeds the low-frequency a-c, while for $\nu>2$ the reverse is true. Contrast this with the behavior of broad-band noise, Fig. 8. We see also in Fig. 10 that over the usual range of ν, i.e., $1<\nu<2$, the power associated with the higher zones ($l\geqq 3$) is quite negligible for most purposes compared with that in the zones 0, 1, and 2. In the instance of narrow-band noise the fundamental ($l=1$) spectral region appears to have the greatest relative mean power. Examples of spectra for regions 0 and 1 are to be found in reference 9.

8. Small-signal detection and full-wave rectification. The method of section 7 is particularly well suited to the study of the important practical case of small-signal detection. This process is characterized by input amplitudes sufficiently small so that cut-off, corresponding to large grid-voltage swings in the negative direction, and saturation, arising from excessively large positive swings, are both avoided. For periodic disturbances or in general for waves limited in amplitude, it is possible to achieve this condition. For noise we must modify the "small-signal" concept by specifying that the instantaneous voltage amplitude of the input does not enter the regions of saturation and cut-off (*vide* Fig. 1) more than a given percentage of the time. In

our analysis we assume that this percentage is small enough to permit us to replace the physically bounded dynamic characteristic by one unaffected by cut-off or saturation.

We may represent the small-signal dynamic path, or dynamic transfer characteristic, as it is sometimes called,[30] by the following

$$I = \alpha_0 + \alpha_1 V + \alpha_2 V^2 + \cdots + \alpha_k V^k + \cdots = \sum_{k=0}^{\infty} \alpha_k V^k, \tag{8.1}$$

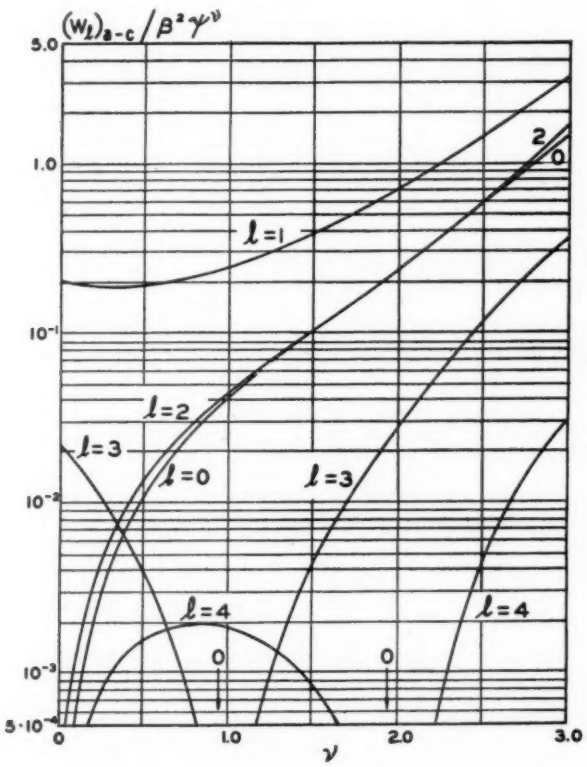


FIG. 10. The mean a-c power outputs associated with the various output harmonic regions ($l=0-4$), as a function of ν.

where V is again identified as the disturbance on the grid and I is the instantaneous current output; the series (8.1) is assumed to converge. The quantities α_k are dimensional constants, for resistive tube loads, and may be described in terms of the tube parameters μ, r_b, r_p, $\partial r_p/\partial e_b$ etc. (See Eqs. 3.41, 3.42, reference 30.) For reactive loads, however, the α_k are complicated functions of the frequency and hence of the input and output waves. As before, our study is restricted to cases where the plate loads are primarily resistive.

[30] H. J. Reich, *Theory and application of electron tubes*, McGraw-Hill (1944) pp. 74–77.

(*a*) *Noise Alone.* When the applied voltage $V(t)$ is random noise, the correlation function of the output is from (2.9) and 1.7b) in normalized form

$$R(t)_N = [2\pi(1-r^2)^{1/2}]^{-1}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dyI(x\psi^{1/2})I(y\psi^{1/2})e^{-(x^2+y^2-2xyr)/2(1-r^2)}. \tag{8.2}$$

The substitution of (8.1) in (8.2) gives

$$R(t)_N = [2\pi(1-r^2)^{1/2}]^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\alpha_j\alpha_k\psi^{(j+k)/2}\int_{-\infty}^{\infty}x^j dx\int_{-\infty}^{\infty}y^k dy e^{-(x^2+y^2-2xyr)/2(1-r^2)}. \tag{8.3}$$

The integral in (8.3) may be evaluated with the help of polar coördinates, the method of section 7, or by expanding exp. $(xyr/(1-r^2))$ and applying (A3.20b). The result in either case is

$$\begin{aligned}
Q_{j,k}(r) &= [2\pi(1-r^2)^{1/2}]^{-1}\int_{-\infty}^{\infty}e^{-x^2/2(1-r^2)}x^j dx\int_{-\infty}^{\infty}y^k e^{-y^2/2(1-r^2)}e^{xyr/(1-r^2)}dy \\
&= 0, \qquad j+k \text{ odd}, \\
&= 2r\cdot\frac{2^{(j+k)/2}}{\pi}\,\Gamma\left(\frac{j}{2}+1\right)\Gamma\left(\frac{k}{2}+1\right) \\
&\qquad \times {}_2F_1\left(\frac{1-j}{2},\ \frac{1-k}{2};\ \frac{3}{2};\ r^2\right), \qquad i,\ k \text{ both odd}, \\
&= \frac{2^{(j+k)/2}}{\pi}\,\Gamma\left(\frac{j+1}{2}\right)\Gamma\left(\frac{k+1}{2}\right) \\
&\qquad \times {}_2F_1\left(-\frac{j}{2},\ -\frac{k}{2};\ \frac{1}{2};\ r^2\right), \qquad j,\ k \text{ both even.}
\end{aligned} \tag{8.4}$$

Observe that the hypergeometric functions terminate. Equation (8.2) may be written finally

$$R(t)_N = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\alpha_j\alpha_k\psi^{(j+k)/2}Q_{j,k}(r), \qquad j+k \text{ even.} \tag{8.5}$$

Only when $j+k$ is even is there a non vanishing output. This quite naturally follows from the basic assumption for our random noise that it contains no *d-c* and consequently has the average value zero, i.e., $\bar{x}=\bar{y}=0$, and so $\overline{x^j}$ and $\overline{y^k}$ also vanish (j, k odd) for the infinite time average. Then it is clear that $\overline{x^j y^k}$ *on the average* must be zero provided $j+k$ is odd, as borne out by (8.4). This also follows at once from (A1.4) et seq. when $\nu_k=0$.

The output spectrum is obtained from (8.5) and (1.6), and the total output power is found as before on setting $t=0$ in $R(t)_N$. With the aid of (A3.20a) we have

$$W_\tau = R(0) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\alpha_j\alpha_k\left(\frac{\psi}{2}\right)^{(j+k)/2}(j+k)!\Big/\left(\frac{j+k}{2}\right)!, \qquad j+k \text{ even.} \tag{8.6}$$

In practice only the first few terms are generally significant, because α_j and α_k ap-

proach zero rapidly for larger values of j and k. The d-c output is easily derived on setting $t=\infty$ in (8.5). The expression is

$$W_{d-c} = R(\infty) = \left\{ \sum_{j=0}^{\infty} \alpha_j \left(\frac{\psi}{2}\right)^{j/2} j! \Big/ \left(\frac{j}{2}\right)! \right\}^2, \qquad j = 0, 2, 4, \cdots . \tag{8.7}$$

Equation (8.7) shows that only the even terms in the dynamic characteristic contribute to the d-c. The mean power associated with the continuum is now to be obtained from (8.6)–(8.7) and the fundamental relation $W_c = W_r - W_{d-c}$.

When the input noise is broadband there is no separation of the output into spectral bands centered about harmonics of the central frequency as explained in section 3. The spectrum must be calculated directly from (8.5). On the other hand, when the noise is narrow-band, a resolution into harmonics is possible, for which the various correlation functions and hence the powers and spectra may be determined. We note again that the input correlation is given by (A2.8).

Let us consider first the low-frequency output, as it is perhaps of greatest practical interest. Now we observe from (8.4) that only when j and k are both even does $r(t)$ exist in even powers and therefore contributes to the d-c and low-frequency continuum. Our procedure for selecting the terms that contribute is similar to that outlined in section 7. Application of the expansion of $(\cos \omega_c t)^{2n}$ to the hypergeometric function in (8.4) yields

$$\begin{aligned} &{}_2F_1\left(-\frac{j}{2}, \ -\frac{k}{2}\ ; \ \frac{1}{2}\ ; \ r^2\right) \\ &\quad = \sum_{n=0}^{j/2,\,k/2} \frac{(-j/2)_n(-k/2)_n r_0^{2n}}{(\frac{1}{2})_n n!\, 2^{2n}} \sum_{q=0}^{n} \epsilon_{n-q}\, {}_{2n}C_q \cos 2(n-q)\omega_c t, \quad j, k \text{ even}, \end{aligned} \tag{8.8}$$

where either limit applies on the first summation, according to whichever is the lesser, $j/2$ or $k/2$. Only those terms for which $q=n$ contribute to the low frequency and d-c. We have then

$$\begin{aligned} &{}_2F_1\left(-\frac{j}{2}, \ -\frac{k}{2}\ ; \ \frac{1}{2}\ ; \ r^2\right)_0 \\ &\quad = \sum_{n=0}^{j/2,\,k/2} \frac{(-j/2)_n(-k/2)_n r_0^{2n}\ {}_{2n}C_n}{(\frac{1}{2})_n n!\, 2^{2n}} = {}_2F_1\left(-\frac{j}{2}, \ -\frac{k}{2}\ ; \ 1; \ r_0^2\right) \quad j, k \text{ even}, \end{aligned} \tag{8.9}$$

since $(1/2)_n = (2n)!/n!2^{2n}$. Equation (8.5) becomes finally

$$\begin{aligned} R_0(l)_N = \frac{1}{\pi} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \alpha_j \alpha_k (2\psi)^{(j+k)/2} \Gamma\left(\frac{j+1}{2}\right) \Gamma\left(\frac{k+1}{2}\right) \\ \times {}_2F_1\left(-\frac{j}{2}, \ -\frac{k}{2}\ ; \ 1; \ r_0^2\right), \qquad j, k \text{ even}. \end{aligned} \tag{8.10}$$

The low-frequency continuum is calculated from (8.10) when all the constant terms in $\sum {}_2F_1$ corresponding to the d-c are removed. The mean total power for the harmonic zone $l=0$ follows from (8.10) and (A3.20a), when $t=0$. The d-c power output is

seen to be precisely (8.7), as we would expect, and the continuum power $(W_c)_0$ follows from $(W_\tau)_0 - W_{d-c} = (W_c)_0$.

The above procedure may be applied to the higher spectral zones $l \geqq 1$ where, following the analysis of Sec. 7(b), we may derive the interesting partial sum results

$$
{}_2F_1\left(-\frac{j}{2},\ -\frac{k}{2};\ \frac{1}{2};\ r_0^2\cos^2 l\omega_c t\right)
$$

$$
= \sum_{l=0,2,\cdots}^{\infty} \epsilon_{l/2}\cos l\omega_c t\,\frac{r_0^l}{l!}(-j/2)_{l/2}(-k/2)_{l/2}
$$

$$
\times\ {}_2F_1\left(\frac{l-j}{2},\ \frac{l-k}{2};\ l+1;\ r_0^2\right), \qquad i,\ k \text{ even}, \tag{8.11}
$$

and

$$
r_0\cos\omega_c t\ {}_2F_1\left(\frac{1-j}{2},\ \frac{1-k}{2};\ \frac{3}{2};\ r_0^2\cos^2\omega_c t\right)
$$

$$
= \sum_{l=1,3,\cdots}^{\infty} \cos l\omega_c t\,\frac{r_0^l}{l!}\left(\frac{1-j}{2}\right)_{(l-1)/2}\left(\frac{1-k}{2}\right)_{(l-1)/2}
$$

$$
\times\ {}_2F_1\left(\frac{l-j}{2},\ \frac{l-k}{2};\ l+1;\ r_0^2\right), \qquad j,\ k \text{ odd}. \tag{8.12}
$$

From (8.11) and (8.12) substituted into (8.5) one may write at once for the correlation functions of the l respective output bands

$$
R_l(t)_N = \epsilon_{l/2}\,\frac{r_0^l\cos l\omega_c t}{l!}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\alpha_j\alpha_k(2\psi)^{(j+k)/2}(-j/2)_{l/2}(-k/2)_{l/2}\Gamma\left(\frac{j+1}{2}\right)\Gamma\left(\frac{k+1}{2}\right)
$$

$$
\times\ {}_2F_1\left(\frac{l-j}{2},\ \frac{l-k}{2};\ l+1;\ r_0^2\right), \qquad j,\ k \text{ even} \tag{8.13}
$$

for even values of l, and when l is odd we obtain

$$
R_l(t)_N = \frac{2r_0^l\cos l\omega_c t}{l!}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\alpha_j\alpha_k(2\psi)^{(j+k)/2}\left(\frac{1-j}{2}\right)_{(l-1)/2}\left(\frac{1-k}{2}\right)_{(l-1)/2}
$$

$$
\times\ \Gamma\left(\frac{j}{2}+1\right)\Gamma\left(\frac{k}{2}+1\right){}_2F_1\left(\frac{l-j}{2},\ \frac{l-k}{2};\ l+1;\ r_0^2\right),\ j,\ k \text{ odd}. \tag{8.14}
$$

From these relations it is observed that for given values of j and k it is not possible to have contributions to zones for which l exceeds j or k, whichever is the lesser. Equation (8.13) and (8.14) also enable us to determine the various mean a-c powers in the continuum of the output, on setting $t=0$ as before, and with the aid of (A3.20a).

(*b*) *Signal and Noise.* Instead of the input voltage being merely random noise, it now consists of a mixture of noise and a signal S, where the latter may or may not be modulated. Since the noise and signal are independent, there can be no correlation between them. Further, if the signal is modulated, correlation may or may not exist between the modulation and the carrier, but in any case there is still no correlation

with the noise; (see the first part of section 2). Accordingly, in determining the correlation function that part, R_N, is obtained by a suitable average over the random noise components independently of the contribution arising from the signal. The complete correlation follows from (2.12).

Let S_1 and S_2 be the input signal voltages at times separated by an interval t. Then the total input voltage may be represented in normalized form by

$$\begin{aligned} v_1 &= V_1/\psi^{1/2} = (S_1 + X)/\psi^{1/2} = s_1 + x; \\ v_2 &= V_2/\psi^{1/2} = (S_2 + Y)/\psi^{1/2} = s_2 + y. \end{aligned} \tag{8.15}$$

The contribution to the correlation, attributable to the noise, follows from (8.2) where now the input is given by (8.15). We find then that

$$R_N = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \alpha_j\alpha_k\psi^{(j+k)/2} \int_{-\infty}^{\infty}(x + s_1)^j dx \int_{-\infty}^{\infty}(y + s_2)^k W_2(x, y; t)dy, \tag{8.16}$$

with $W_2(x, y; t)$ given by (2.9) after suitable normalization. We note several methods of evaluating these integrals. Unfortunately, none of these methods yields very simple results, although they appear to be the best available. However, in practice the dynamic characteristic (8.1) may be expressed with reasonable accuracy when only the first few terms are considered; the higher coefficients α_j, α_k, $j, k \geqq 3$ are in many cases negligible, and the complexity of the results is consequently much reduced. One approach is that used in evaluating (8.3). Applying it to (8.16) one obtains for the coefficient of $\alpha_j\alpha_k\psi^{(j+k)/2}$:

$$\begin{aligned} I_{j,k} &= \int\int_{-\infty}^{\infty}(x + s_1)^j(y + s_2)^k W_2(x, y; t)dxdy \\ &= j!k!2^{-(j+k)/2} \sum_{p=0}^{j}\sum_{q=0}^{k} \frac{2^{(p+q)/2}s_1^p s_2^q}{p!q![(j-p)/2]![(k-q)/2]!} \\ &\times {}_2F_1\left(\frac{p-j}{2}, \frac{q-k}{2}; \frac{1}{2}; r^2\right), \qquad j-p,\ k-q \text{ even}, \\ &= 2rj!k!2^{-(j+k)/2} \sum_{p=0}^{j}\sum_{q=0}^{k} \frac{2^{(p+q)/2}s_1^p s_2^q}{p!q![(j-p-1)/2]![(k-q-1)/2]!} \\ &\times {}_2F_1\left(\frac{1-j+p}{2}, \frac{1-k+q}{2}; \frac{3}{2}; r^2\right), \quad j-p,\ k-q \text{ odd}, \end{aligned} \tag{8.17}$$

and when $j+k-(p+q)(\geqq 0)$ is odd the integrals vanish. We may also derive (8.17) with the aid of the transformation to polar coördinates employed in section 7.

Now let us assume that the modulation A_0 and the carrier $\cos \omega_0 t$ are uncorrelated. Then the input signal S may be written

$$S_1(t_0, t_0') = A_0(t_0') \cos \omega_0 t_0, \qquad S_2(t_0 + t, t_0' + t) = A_0(t_0' + t) \cos \omega_0(t_0 + t). \tag{8.18}$$

The complete correlation may be found from (2.12) on averaging over the phases t_0 and t_0', and from (2.12) the mean power spectrum follows at once, the mean power on setting $t=0$. As an example, consider the general small-signal quadratic rectifier α_j, $\alpha_k \neq 0$, $j \leqq 2$, $k \leqq 2$. With the help of (8.17)[31] we find for R_N

[31] See Part II of reference 8, section II b.

$$R_N = \alpha_0^2 + \psi^{1/2}\alpha_0\alpha_1(s_1 + s_2) + \psi\alpha_0\alpha_2(2 + s_1^2 + s_2^2) + \psi\alpha_1^2(r + s_1s_2)$$
$$+ \psi^{3/2}\alpha_1\alpha_2(s_1 + s_2)(s_1s_2 + 2r + 1)$$
$$+ \psi^2\alpha_2^2(1 + 2r^2 + s_1^2 + s_2^2 + 4rs_1s_2 + s_1^2s_2^2). \tag{8.19}$$

If the noise is narrow-band, $r = r_0 \cos \omega_c t$ and further, if the carrier is tuned to the center of the noise band, as is the case in receivers or transmitters, then $\omega_c = \omega_0$. Thus, when (8.19) is substituted into the expression for the complete correlation function (2.12) and when (8.18) is used, we observe that the terms involving s_1, s_2, s_1^2, s_2^2, $s_1^2s_2$, $s_1s_2^2$, s_1s_2, s_1r, s_2r, but *not* s_1s_2r, contribute only to the *d-c* and to frequencies in the neighborhood of f_c and $2f_c$. The low-frequency output, exclusive of direct current, is from (8.19) in conjunction with (2.12).

$$R(t)_{L\cdot F.} = \psi^2\alpha_2^2\left\{r_0(t)^2 + r_0(t)T_0'^{-1}\int_0^{T'}(A_{01}A_{02})_{L\cdot F.}dt_0'\right.$$
$$\left. + (4T_0')^{-1}\int_0^{T'}(A_{01}^2A_{02}^2)_{L\cdot F.}dt_0'\right\}, \tag{8.20}$$

where we have written A_{01} (for $A_0(t_0')$ and A_{02} for $A_0(t_0' + t)$, and the subscript $L - F$ indicates that *d-c* components are to be removed. Observe that the low-frequency output is composed of three contributions: the first in (8.20) is that arising from the input noise alone, the second represents the cross-modulation of the noise and signal components, and the third is the detected signal envelope, squared, of course, since in this instance the dynamic characteristic is quadratic. We remark that only the term in α_2^2, cf. (8.1), is capable of rectifying; hence the low-frequency output alone results from it. Special cases of (8.19) and (8.20) have been given by Rice, ref. 2, Eqs. (4.10.1) and (4.10.3), when the incoming wave is noise alone, or consists of noise and an unmodulated carrier.

As before, *vide* (8.6) and (8.7), it is possible to deduce (from 8.16) and (8.17) certain general relations for the output, total continuum, and *d-c* power.

(*c*) *Full-Wave Rectification.* By full-wave rectification it is meant that the dynamic characteristic is such that $I_{\text{out}} = \beta|V_{\text{in}}|^r$. When the incoming wave is noise alone, we have merely to multiply the results obtained in section 7 by 2^r. However, if the input contains a signal as well, Eq. (7.2) multiplied by 2^r still applies, provided we write for the lower limits of integration $-s_1$ and $-s_2$ respectively, in place of $x = y = 0$. Analytically the problem is now best handled by the methods of section 2 and 3.

We wish to thank Mr. Rice and Prof. J. H. Van Vleck with whom we have discussed these problems from time to time, and also Prof. L. Brillouin, for their helpful criticism of this paper.

Appendix I. In this appendix are briefly summarized the analytical results obtained when the solution to the generalized, *s*-dimensional problem of random flights is used, with the aid of the characteristic function and central limit theorem,[10,32] to derive the multivariate Gaussian distribution law, so fundamental in all problems of the type considered in this paper.

The generalized random-walk problem,[4] resolved by a method due originally to

[22] J. V. Uspensky, *Introduction to mathematical probability*, McGraw-Hill (1937).

Markoff,[33] leads to an expression for the characteristic function $F_N(\xi)$, whose associated probability density $W_N(\mathbf{R})$ it is our task to find: viz.

$$F_N(\xi) = \exp\left\{i\sum_{k=1}^{s}\nu_k\xi_k - \frac{1}{2}\sum_{k=1}^{s}\sum_{l=1}^{s}\mu_{kl}\xi_k\xi_l\right\}, \qquad N \gg 1, \tag{A1.1}$$

where the quantities ν_k and μ_{kl} are given by

$$\nu_k = \sum_{j=1}^{N}\bar{x}_{jk}; \qquad \mu_{kl} = \sum_{j=1}^{N}\overline{x_{jk}x_{jl}}, \qquad N \gg 1. \tag{A1.2}$$

Here $\mathbf{r}_j$ is the jth displacement vector, with components $r_{jk}=x_{jk}$ $(k=1, 2, \cdots, s)$, and the range of values of x_j extends from $-\infty$ to $+\infty$; N is the total number of displacements, here indefinitely large, and $\mathbf{R}$ is the resultant displacement. Since F_N is the Fourier transform of W_N, we have finally, by the usual methods,[34] the familiar result that[35]

$$W_N(R) = \frac{\exp -\frac{1}{2}(\tilde{\mathbf{X}}-\tilde{\nu})\mathbf{M}(\mathbf{X}-\nu)}{(2\pi)^{s/2}\,|\,\mu\,|^{1/2}}$$

$$= [(2\pi)^s\,|\,\mu\,|\,]^{-1/2}\exp\left\{-\frac{1}{2}\sum_{k=1}^{s}\sum_{l=1}^{s}\frac{\mu^{kl}}{|\,\mu\,|}(X_k-\nu_k)(X_l-\nu_l)\right\}, \tag{A1.3}$$

where X_k is the kth component of $\mathbf{R}$ and $\overline{X}_k=\nu_k$; here $\mathbf{M}$ is an (sxs) matrix reciprocal to the matrix μ: $|\mu|$ is the determinant of μ, and μ^{kl} is the cofactor of μ_{kl}. Since the average of $\exp(i\sum_k\xi_kX_k)$ is, by definition, the characteristic function, it is a simple matter to show that

$$\overline{X}_k = \nu_k = \sum_{j=1}^{N}\bar{x}_{jk}; \quad \overline{X_{k_1}X_{k_2}} = \mu_{k_1}\mu_{k_2} + \nu_{k_1}\nu_{k_2} = \sum_{j=1}^{N}(\overline{x_{jk_1}x_{jk_2}} + \overline{x_{jk_1}}\cdot\overline{x_{jk_2}}) \tag{A1.4}$$

by expanding this and (A1.1) and equating coefficients of ξ_k, $\xi_{k_1}\xi_{k_2}$, etc. In the same way higher moments may be found, it being noted that if $\nu_k=0$, all odd-order moments vanish.

Appendix II: Some Remarks on Correlation Functions, Spectra, and Power. The fundamental relationship between the mean power—or mean square amplitude spectrum—and the correlation function, defined below, is well known. We mention it briefly here along with a short discussion of some of the more significant and useful properties of the relation in our work, some of which do not seem to have been treated previously. There appear to be several approaches, yielding similar results, one, for example, through the use of Fourier series,[1] the other with the help of Plancherel's theorem.[36]

Now let us consider $g(t)$ to represent a suitable function of the time, and let us

---

[33] A. Markoff, *Wahrscheinlichkeitsrechnung* (Leipzig, 1912).

[34] See ref. 3; also sec. 10:16 of Margenau and Murphy, *The mathematics of physics and chemistry* (D. Van Nostrand, 1943) and ref. 2, sec. 3.5.

[35] $\mathbf{X}$ and ν are column matrices, and $\sim$ indicates the transposed matrix.

[36] See E. C. Kemble, *Fundamental principles of quantum mechanics* (McGraw-Hill, p. 36, 1937); also M. Plancherel, Rend. di Palermo **30**, 289 (1910). The theorem is also known by Parseval's name.

require that $g(t)$ be in general different from zero in a long time interval T, but be zero outside this interval. With this in mind we may consider the *correlation function* for $g(t)$, which is defined as

$$R(t) \equiv \lim_{T\to\infty} \frac{1}{T}\int_{-\infty}^{\infty} \overline{g(t_0)g(t_0+t)}dt_0 = \overline{[g(t_0)g(t_0+t)]}_{\text{av.}}, \tag{A2.1}$$

where now $g(t_0)$ vanishes outside the interval $0<t<T$. The average in (A2.1) is to be performed over all phases of the disturbance, and the bar indicates the average computed over any random variables in the wave: for it often happens that $g(t)$ is a stochastic or random function of the time t, or at least that some component of $g(t)$ is randomly distributed. Analytically such randomness is introduced by treating the function as involving a certain number of parameters, and then taking these parameters to be random variables, distributed according to a certain law. Now for any given set of parameters there will be a definite correlation function $R(t)$,and as we shall see, a definite mean-square amplitude spectrum also. However, it is the *average* of this set of correlation functions and spectra that is important for our work. These averages may be obtained by averaging over the ranges of the parameters, with the help of their distributions, as indicated by the superscript bar on (A2.1) and elsewhere. We remark further that should $g(t)$ comprise two or more incommensurable periodic disturbances, the definition (A2.1) is easily extended to include the separate averages over the respective phases of the additional components, inasmuch as the correlation function of such a wave is simply the product of the correlation functions of the separate parts. An important example of this type of function is the modulated carrier, *vide* Sec. 2, where no correlation between carrier and modulation exists.

Letting

$$g_1 = g(t_0), \quad (0 < t_0 < T); \qquad g_2^* = g(-t_0 + t), \quad (t_0 - T < t < t_0), \tag{A2.2}$$

and with the help of the Fourier transforms of g_1 and g_2^* and Plancherel's theorem, we obtain finally for (A2.1) the correlation function

$$R(t) = \lim_{T\to\infty} \int_{-\infty}^{\infty} \frac{\overline{|S(f)|^2}}{T} e^{-i\omega t}df = \int_0^{\infty} W(f)\cos\omega t df, \qquad f>0, \tag{A2.3}$$

where $W(f)$ is the mean spectral density defined as

$$W(f) = \lim_{T\to\infty} \frac{\overline{2|S(f)|^2}}{T}, \qquad f>0, \tag{A2.4}$$

and $S(f)$ is the amplitude spectrum of $g(t)$. The inversion of (A2.3) gives us at once the other relation

$$W(f) = 4\int_0^{\infty} R(t)\cos\omega t dt. \tag{A2.5}$$

The limit in (A2.4) is assumed exist. When it does not, the physical significance of (A2.4) is this: it represents the power at some discrete frequency f_0 rather than a spectral distribution over a continuous range of frequencies. Then analytically in the

limit $W(f)$ is observed to become infinite at $f=f_0$ and be zero elsewhere, such that $\int_0^\infty W(f)df$ converges. The spectral density $W(f)$ accordingly exhibits the properties of a delta-function.

On setting $t=0$ we find from (A2.1) and (A2.3) that

$$R(0) = \lim_{T\to\infty} \int_{-\infty}^{\infty} \frac{\overline{g(t_0)^2}}{T}\, dt_0 = [\overline{g(t_0)^2}]_{\text{av.}} = \int_0^\infty W(f)df, \qquad \text{(A2.6)}$$

showing that the mean total power may be obtained from the correlation function by putting t equal to zero in the latter; this is in agreement with our definition (A2.4) of the spectral density. Furthermore, *the mean total power is observed* from (A2.6) *to be independent of the shape of the spectrum*, depending only on the integral $\int_0^\infty W(f)df$. It should also be mentioned that the spectrum (A2.4) can never give us the time-variation of the wave, since information about the phases is always wanting. Hence an infinite number of different functions $g(t)$ may be combined to give the same value of $W(f)$.

It is interesting to observe what happens after very long times. For a purely stochastic disturbance, which from (A2.1) is seen to be independent of the average over T, $R(\infty)$ becomes zero: there ceases to be any correlation at all between an event at time t_0 and one at t_0+t, $t>\infty$ later. But for periodic components $R(\infty)$ approaches no definite limit, since this part of the disturbance is indefinitely repeated and can never be said ultimately to die down to zero in time. By considering the oscillatory or constant parts of the expression for the correlation function in the limit $t\to\infty$, we can determine the contribution to the total power arising from the periodic part of the wave, for with the help of (A2.5) and the delta-function the result is seen to be (the sum of) the mean powers in the respective components. These quantities may also be identified as the coefficients of the constant or of the trigonometric parts of the correlation function $R(t)$. The constant part corresponds to d-c, the others, to the various discrete frequencies. Thus, in turn, the power in the continuum is $R(0)-R(\infty)$, and may be obtained from that portion of the correlation function which vanishes at $t=\infty$. Examples are given in Sections 2 and 7.

It is convenient to use the *normalized* correlation function $r(t)$, where

$$r(t) \equiv \psi(t)/\psi(0) = \int_0^\infty w(f)\cos\omega t df \Big/ \int_0^\infty w(f)df; \qquad \rho(t) \equiv R(t)/R(0). \qquad \text{(A2.7)}$$

Here $\psi(t)$ is chosen to represent the correlation of a wave entering some non linear device and $w(f)$ is its mean power spectrum; $\rho(t)$, correspondingly, is the normalized correlation of the output, when $R(t)$ is so distinguished. It follows from (A2.7) that $\rho(0)=r(0)=1$, which is the maximum value of $\rho(t)$ and $r(t)$.

The important special case in which the incoming noise is confined to a symmetrical band of frequencies narrow compared with the central frequency f_c, i.e., $w=w(f-f_c)$ (see Fig. 7c), leads to

$$r(t) = \psi(0)^{-1}\int_0^\infty w(f-f_c)\cos\omega t df \doteq \left\{\psi(0)^{-1}\int_{-\infty}^{\infty} w(f')\cos\omega' t df'\right\}\cos\omega_c t$$
$$\equiv r_0(t)\cos\omega_c t, \qquad \text{(A2.8)}$$

on the change of variable $f-f_c=f'$, where we have ignored the spectral "tail" at $f=0$, inasmuch as $w(-f_c)$ is assumed to be very much smaller than $w(0)$. As an example, consider the Gaussian spectrum $w(f)=W_0\exp[-(f-f_c)^2/f_b^2]$, $f_c/f_b\gg 1$; for this $\psi(t)$ and $\psi(0)$ are readily shown to be

$$\psi(t)=\psi(0)r_0(t)\cos\omega_c t=\left(\frac{W_0\omega_b}{2\pi^{1/2}}\right)e^{-\omega^2 t_b^2/4}\cos\omega_c t;\quad \psi(0)=\frac{W_0\omega_b}{2\pi^{1/2}}; \tag{A2.9}$$

W_0 is the maximum spectral density. In a similar fashion correlation functions for other input spectral distributions are easily determined with the help of (A2.7) or (A2.8).

One may expand the output correlation function $R(t)$ in a power series in $\psi(t)$ (or $r(t)$), the input correlation, and hence as a function of $\cos l\omega_c t$ ($l=0, 1, 2, \cdots$), for the case of narrow-band noise, by virtue of (A2.8) and the resolution of $(\cos\omega_c t)^n$ into harmonics.[9] The series then becomes

$$R(t)=\sum_{l=0}^{\infty}G_l(t)\cos l\omega_c t=\sum_{l=0}^{\infty}R_l(t), \tag{A2.10}$$

where now $R_l(t)$ is the correlation of the lth harmonic zone generated in the output. The spectrum of these bands follows at once from (A2.5), showing also that the various "resolved" spectra $W_l(f)$ are distributed about the harmonics lf_c ($l=0, 1, 2\cdots$). Examination of one such region $W_l(f)$ shows that here, too as in (A2.6), the mean power in any given band is independent of the spectral distribution of the band. Further, since the input spectrum enters only through $\psi(t)$ and $\psi(0)$ (= the mean input power) *the mean power in the l-th band is also independent of the original spectral shape w(f)* of the incoming wave. This is, strictly speaking, only practically true, not rigorously so, as distortion of the input always spreads its spectrum and thus spectral "tails" from one region overlap those of another. This overlapping is quite insignificant most of the time, as long as the band is narrow, the criterion of narrowness depending on what is considered a negligible spectral ordinate.

Appendix III: Special Functions and Integrals. A quantity that appears often in our analysis, cf. Secs. 4 or 5, is the confluent hypergeometric function

$${}_1F_1(\alpha;\beta;x)=1+\frac{\alpha x}{\beta 1!}+\frac{\alpha(\alpha+1)}{\beta(\beta+1)}\frac{x^2}{2!}+\cdots\frac{(\alpha)_n x^n}{(\beta)_n n!}+\cdots, \tag{A3.1}$$

where $(\alpha)_n=\alpha(\alpha+1)\cdots(\alpha+n-1)$, and $(\alpha)_0=1$, as usual. This function has the important property, known as Kummer's transformation,[21] that

$${}_1F_1(\alpha;\beta;x)=e^x\,{}_1F_1(\beta-\alpha;\beta;-x), \tag{A3.2}$$

and it may be shown that the asymptotic development of ${}_1F_1$ takes the form,[37]

$${}_1F_1(\alpha;\beta;-x)\simeq\frac{\Gamma(\beta)}{\Gamma(\beta-\alpha)}x^{-\alpha}\left\{1+\frac{\alpha(\alpha-\beta+1)}{x1!}+\frac{\alpha(\alpha+1)(\alpha-\beta+1)(\alpha-\beta+2)}{x^2 2!}+\cdots\right\},\qquad R(x)>0. \tag{A3.3}$$

[37] See, for example, Whittaker and Watson, *Modern analysis* (Cambridge Univ. Press, 1940), Chapter XVI.

The expression (A3.3) is useful in determining the limiting forms of spectra and power distribution when the root-mean-square noise voltage $\psi^{1/2}$ is much less than either the cut-off voltage b_0 or the amplitude of the carrier; see Sec. 4. The confluent hypergeometric functions of both negative and positive argument may also be expressed in terms of the modified Bessel functions of the first kind, for certain combinations of values of α and β. Examples are given in Appendix II of Bennett's paper[6] and also in ref. 2.

Another function of considerable interest is

$$\phi^{(j)}(b) = \frac{d^{(j)}}{dx^j}\left.\frac{e^{-x^2/2}}{\sqrt{2\pi}}\right|_{x=b} \equiv (-1)^j H_j(b) e^{-b^2/2}/(2\pi)^{1/2}, \qquad j = 0, 1, 2, \cdots . \tag{A3.4}$$

Here the $H_j(b)$ are Hermitian polynomials[38] of order j. It is not difficult to show, by successive differentiations with respect to b_0 when $n=0$, that the ϕ-functions may be given as the following infinite integrals

$$\begin{aligned}
\int_0^\infty z^{2n} \cos b_0 z e^{-\psi z^2/2} dz &= \left(\frac{\pi}{2}\right)^{1/2} (-1)^n \psi^{-n-1/2} e^{-b_0^2/2\psi} H_{2n}(b_0/\psi^{1/2}) \\
&= \pi(-1)^n \psi^{-n-1/2} \phi^{(2n)}(b_0/\psi^{1/2}) \\
\int_0^\infty z^{2n-1} \sin b_0 z e^{-\psi z^2/2} dz &= \left(\frac{\pi}{2}\right)^{1/2} (-1)^{n+1} \psi^{-n} e^{-b_0^2/2\psi} H_{2n-1}(b_0/\psi^{1/2}) \\
&= \pi(-1)^n \psi^{-n} \phi^{(2n-1)}(b_0/\psi^{1/2}), \quad n = 0, 1, 2, \cdots ,
\end{aligned} \tag{A3.5}$$

and in this connection

$$\phi^{(-1)}(b_0/\psi^{1/2}) = \frac{1}{\sqrt{2\pi}} e^{-b_0^2/2\psi} H_{-1}(b_0/\psi^{1/2}) = \tfrac{1}{2}\Theta(b_0/\sqrt{2\psi}), \text{ where } \Theta(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-y^2} dy$$

is the familiar error function, tabulated, for example, in Jahnke and Emde and in Pierce's Tables. Tables of $\phi^{(n)}$ for $n=0, \cdots, 6$ are given in T. C. Fry's *Probability and its engineering uses* (D. Van Nostrand, 1928, pp. 456, 457). Additional values may be obtained from the recurrence relation

$$\phi^{(n+1)}(b) = -\left\{b\phi^{(n)}(b) + n\phi^{(n-1)}(b)\right\}, \qquad n = 0, 1, 2, \cdots . \tag{A3.6}$$

We may also express $\phi^{(n)}$ in terms of the confluent hypergeometric function ${}_1F_1(\alpha;\beta;-x)$. To do this we need Hankel's exponential integral[39]

$$\int_0^\infty J_\nu(az) z^{\mu-1} e^{-q^2 z^2} dz = \frac{\Gamma[(\nu+\mu)/2]}{2q^\mu \Gamma(\nu+1)} \left(\frac{a}{2q}\right)^\nu {}_1F_1\left(\frac{\nu+\mu}{2}; \nu+1; -\frac{a^2}{4q^2}\right),$$
$$R(\mu+\nu) > 0, \qquad |\arg q| < \pi/4, \tag{A3.7}$$

which is readily established by expanding the Bessel function and integrating termwise with help of the Γ-function. Since

$$e^{\pm ib_0 z} = (\pi b_0 z/2)^{1/2}\left\{J_{-1/2}(b_0 z) \pm iJ_{1/2}(b_0 z)\right\}, \tag{A3.8}$$

---

[38] W. Kapteyn, Proc. Royal Acad. Amster. **16,** 1191 (1914), and G. Szegö, Amer. Math. Soc. Colloq. Pub. **23,** 101–104 (1939).

[39] Reference 21, **13.3.**

it follows at once for (A3.6) when this is substituted into (A3.7) that

$$\left.\begin{aligned}\phi^{(2n)}(b_0/\psi^{1/2}) &= \frac{(-1)^n(2n)!}{2^n n!(2\pi)^{1/2}}\,{}_1F_1\left(\frac{2n+1}{2};\ \frac{1}{2};\ -\frac{b_0^2}{2\psi}\right),\\ \text{and}\qquad \phi^{(2n-1)}(b_0/\psi^{1/2}) &= \frac{(-1)^n(2n)!}{2^n n!(2\pi)^{1/2}}\,\frac{b_0}{\psi^{1/2}}\,{}_1F_1\left(\frac{2n+1}{2};\ \frac{3}{2};\ -\frac{b_0^2}{2\psi}\right),\\ &\qquad\qquad n=0,\,1,\,2,\,\cdots.\end{aligned}\right\}\tag{A3.9}$$

We notice also from (A3.1), (A3.4), and (A3.9) that the Hermitian polynomials may be written

$$\left.\begin{aligned}H_{2n}(b) &= \frac{(-1)^n(2n)!}{2^n n!}\,{}_1F_1\left(-n;\ \frac{1}{2};\ \frac{b^2}{2}\right);\\ H_{2n-1}(b) &= \frac{(-1)^{n+1}(2n)!}{2^n n!}\,b\,{}_1F_1\left(1-n;\ \frac{3}{2};\ \frac{b^2}{2}\right),\qquad n=0,\,1,\,2,\,\cdots.\end{aligned}\right\}\tag{A3.10}$$

From these results it is evident that

$$\left.\begin{aligned}\phi^{(2n)}(0) &= \frac{(-1)^n(2n)!}{2^n n!(2\pi)^{1/2}} = H_{2n}(0)/(2\pi)^{1/2};\\ (2\pi)^{1/2}\phi^{(2n-1)}(0) &= -H_{2n-1}(0) = 0.\end{aligned}\right\}\tag{A3.11}$$

Now in the theory of the νth-law non linear device, when $\nu(>0)$ is not necessarily an integer, the integrand of the fundamentals integrals (2.20) contains a branch point at the origin, rather than a simple pole when ν is integral. Accordingly, results like (A3.5) and (A3.7) must be extended to include these more general cases. The first integral to be established is

$$I_1 = \int_{\mathbf{C}} e^{-c^2z^2}z^{2\mu-1}dz = \frac{i\pi e^{-\mu\pi i}}{c^{2\mu}\Gamma(1-\mu)} = ic^{-2\mu}\Gamma(\mu)e^{-\pi\mu i}\sin\pi\mu,\quad |\arg c| < \pi/4,\tag{A3.12}$$

where the contour **C** is the usual one of Eq. (1.8), extending from $-\infty$ to $+\infty$ along the real axis and is indented downward in an infinitesimal semi-circle about $z=0$. Here the argument of z is zero on the positive portion of **C** and is $-\pi$ on the negative part. In the neighborhood of the origin the contribution over the semicircle vanishes in the limit, provided that for the moment we require $R(\mu)>0$. Then we may set $z=e^{-\pi i}t$, $0\leqq t\leqq\infty$, $z<0$; $z=t$, $z>0$, so that finally

$$I_1 = (1-e^{-2\pi\mu i})\int_0^\infty t^{2\mu-1}e^{-c^2t^2}dt = (1-e^{-2\pi\mu i})\Gamma(\mu)/2c^{2\mu} = \pi i e^{-\pi\mu i}/c^{2\mu}\Gamma(1-\mu),$$

$$R(\mu)>0,\qquad |\arg c| < \pi/4.\tag{A3.13}$$

Now $I_1(\mu)$ is certainly analytic for all values of μ, and therefore by analytic continuation we may extend the domain of I_1 to include all values of μ for which $I_1(\mu)$ is finite. This removes the restriction $R(\mu)>0$ and gives us (A3.12).

The generalization of (A3.7) follows from (A3.12). We have only to expand the Bessel function to obtain

$$I_2 = \int_C z^{\mu-1} J_\nu(az) e^{-q^2 z^2} dz = \left(\frac{a}{2}\right)^\nu \sum_{n=0}^{\infty} \frac{(a/2)^{2n}(-1)^n}{n!\Gamma(\nu+n+1)} \int_C z^{\nu+2n-\mu-1} e^{-q^2 z^2} dz, \qquad \text{(A3.14)}$$

termwise integration being allowed because of the absolute convergence of $J_\nu(az)$. Application of (A3.12) with the aid of

$$\Gamma\left(1-n-\frac{\nu}{2}-\frac{\mu}{2}\right) = \Gamma\left(1-\frac{\nu+\mu}{2}\right) \Big/ (-1)^n \left(\frac{\nu+\mu}{2}\right)_n$$

gives us finally

$$I_2 = \frac{\pi i^{1-\nu-\mu}(a/2q)^\nu}{q^\mu \Gamma(\nu+1)\Gamma[1-(\nu+\mu)/2]} {}_1F_1\left(\frac{\mu+\nu}{2};\ 1+\nu;\ -\frac{a^2}{4q^2}\right), \qquad |\arg q| < \pi/4. \qquad \text{(A3.15)}$$

Equation (A3.15) enables us to evaluate the generalized version of (A3.5), namely,

$$I_3 = \int_C z^\mu e^{+izb_0 - c^2 z^2} dz, \qquad |\arg c| < \pi/4. \qquad \text{(A3.16)}$$

The substitution of (A3.8) for $e^{\pm izb_0}$ and application of I_2 gives

$$I_3 = \frac{\pi i^{-\mu}}{c^{\mu+1}} \left\{ {}_1F_1\left(\frac{\mu+1}{2};\ \frac{1}{2};\ -x^2\right) \Big/ \Gamma\left(\frac{1-\mu}{2}\right) \pm 2x\, {}_1F_1\left(\frac{\mu+2}{2};\ -\frac{3}{2};\ -x^2\right) \Big/ \Gamma\left(-\frac{\mu}{2}\right) \right\}, \qquad \text{(A3.17)}$$

where $x = b_0/2c$ and $|\arg.\ c| < \pi/4$, for convergence at infinity. Note that only when μ is an integer can I_3 be expressed as a ϕ-function, cf., Eq. (A3.9).

In the limit of vanishing noise voltage, $\psi \to 0$, or infinite signal or bias and finite noise power, it is necessary to extend the Weber-Schafheitlin integral and evaluate

$$I_4 = \int_C J_\alpha(az) J_\beta(bz) dz/z^\gamma. \qquad \text{(A3.18)}$$

Again, as in the case of the generalized Γ-function, Eq. (A3.12), analytic continuation can be used in precisely similar fashion to give us finally[40]

$$I_4 = \frac{\pi e^{-\pi i(\alpha+\beta-\gamma)/2} b^\beta\, {}_2F_1\left[\frac{1}{2}(\alpha+\beta-\gamma+1),\ \frac{1}{2}(\beta-\gamma-\alpha+1);\ \beta+1;\ b^2/a^2\right]}{2^{\gamma-1} a^{\beta-\gamma+1} \Gamma(\beta+1)\Gamma[(1+\gamma-\alpha-\beta)/2]\Gamma[(1+\gamma+\alpha-\beta)/2]}, \quad 0 \leqq b \leqq a, \qquad \text{(A3.19)}$$

subject only to the restrictions that $R(\gamma) > 0$. The function ${}_2F_1$ has two useful properties, needed in the present work, which we list below:

[40] Reference 21, Sec. **13.4** gives the value of $\int_0^\infty J_\alpha(az) J_\beta(bz) dz/z^\gamma$.

$$ {}_2F_1(\alpha, \beta; \gamma; 1) = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)}, \quad \begin{matrix} R(\gamma) \neq 0, -1, -2, \cdots \\ R(\gamma - \alpha - \beta) \neq 0, -1, -2, \cdots, \end{matrix} \tag{A3.20a} $$

and

$$ {}_2F_1(\alpha, \beta; \gamma; x) = (1 - x)^{\gamma - \alpha - \beta}\, {}_2F_1(\gamma - \beta, \gamma - \alpha; \gamma; x). \tag{A3.20b} $$

(See Chapter XIV of reference 37 for a general treatment.)

NOTES

A NEW DERIVATION OF THE METHOD OF CHARACTERISTICS FOR AXIALLY SYMMETRICAL SUPERSONIC FLOW*

By ANDREW VAZSONYI (*North American Aviation, Inc.*)

The flow of supersonic fluids can be studied theoretically with the aid of the so-called method of characteristics. This method is well known for two-dimensional flows[1] and was extended recently to axially symmetrical flows.[2] The basis of this method

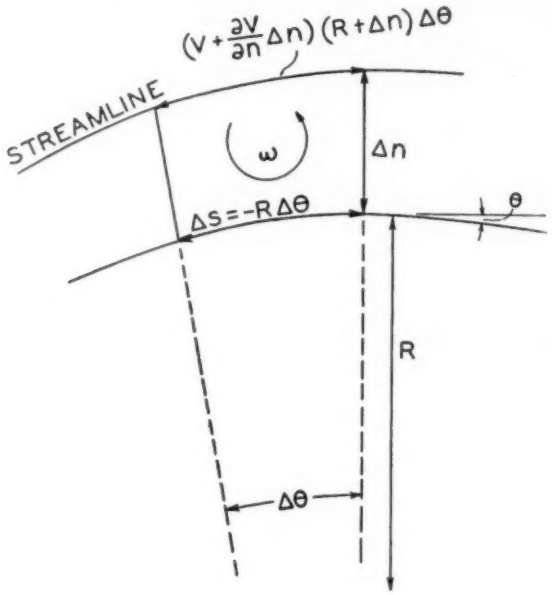


FIG. 1

is a differential equation [Eq. (17) of this paper] expressing the change in the magnitude and the direction of the velocity vector along a Mach wave. The purpose of this note is to give a simple derivation of this equation. Two-dimensional flows and irrotational flows can be obtained by specializing Eq. (17).

1. Bernoulli's equation. If V denotes the velocity, p the pressure, ρ the density, and s the arc length of the stream line, Bernoulli's equation can be written in the form

$$V\frac{\partial V}{\partial s}+\frac{1}{\rho}\frac{\partial p}{\partial s}=0. \tag{1}$$

* Received June 18, 1947.

[1] *Cf.* H. W. Liepman and A. E. Puckett, *Introduction to aerodynamics of a compressible fluid*, John Wiley & Sons, Inc., 1947, pp. 210–232.

[2] *Cf.* R. Sauer, *Theoretische Einführung in die Gasdynamik*, Springer, Berlin, 1943; reprinted by J. W. Edwards, Ann Arbor, 1945, pp. 137–144.

Introducing the velocity of sound defined by

$$a^2 = \frac{dp}{d\rho} \tag{2}$$

we may write this as follows:

$$V\frac{\partial V}{\partial s} - a^2\frac{1}{\rho}\frac{\partial p}{\partial s}. \tag{3}$$

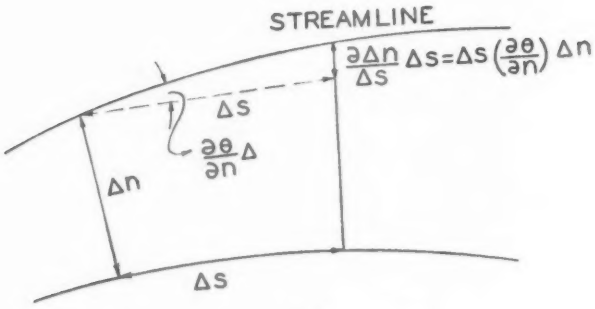


FIG. 2

2. The expression for the rotation. The rotation can be obtained with the aid of Fig. 1 as follows. The circulation around the small element is given by

$$\Gamma = 2\omega\Delta n\Delta s = \left(V + \frac{\partial V}{\partial n}\Delta n\right)(R + \Delta n)\Delta\theta - VR\Delta\theta. \tag{4}$$

Thus,

$$\frac{\partial V}{\partial n} - V\frac{\partial \theta}{\partial s} = -2\omega \tag{5}$$

where R is the radius of curvature of the streamline.

3. The continuity equation. This can be written as

$$2\pi r V\rho\Delta n = \text{const}, \tag{6}$$

where the factor $2\pi r$ enter when the flow is axially symmetrical. By logarithmic differentiation of Eq. (6) we obtain

$$\frac{1}{\rho}\frac{\partial p}{\partial s} + \frac{1}{V}\frac{\partial V}{\partial s} + \frac{1}{\Delta n}\frac{\partial \Delta n}{\partial s} + \frac{\sin\theta}{r} = 0, \tag{7}$$

where

$$\frac{\partial r}{\partial s} = \sin\theta \tag{8}$$

was used. (For two-dimensional flows $r = \infty$.) From Eq. (3) it follows that

$$\frac{1}{\rho}\frac{\partial p}{\partial s}+\frac{1}{V}\frac{\partial V}{\partial s}=\left(1-\frac{V^2}{a^2}\right)\frac{1}{V}\frac{\partial V}{\partial s}=(1-M^2)\frac{1}{V}\frac{\partial V}{\partial s}, \tag{9}$$

where M is the Mach number. From Fig. 2 it can be seen that

$$\frac{1}{\Delta n}\frac{\partial \Delta n}{\partial s}=\frac{\partial \theta}{\partial n}. \tag{10}$$

From Eqs. (7), (9), and (10) follows finally that[3]

$$(M^2-1)\frac{1}{V}\frac{\partial V}{\partial s}-\frac{\partial \theta}{\partial n}-\frac{\sin\theta}{r}=0. \tag{11}$$

4. Let us compute the infinitesimal change in V and θ when moving from point A to point B on Fig. 3. Obviously,

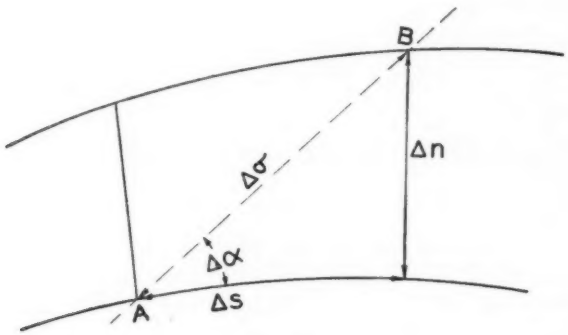


FIG. 3

$$\Delta V=\frac{\partial V}{\partial s}\Delta s+\frac{\partial V}{\partial n}\Delta n=\left(\frac{\partial V}{\partial s}\cos\alpha+\frac{\partial V}{\partial n}\sin\alpha\right)\Delta\sigma \tag{12}$$

and

$$\Delta\theta=\frac{\partial\theta}{\partial s}\Delta s+\frac{\partial\theta}{\partial n}\Delta n=\left(\frac{\partial\theta}{\partial s}\cos\alpha+\frac{\partial\theta}{\partial n}\sin\alpha\right)\Delta\sigma. \tag{13}$$

Using Eqs. (5) and (11), we may transform the last equation into

$$\Delta\theta=\frac{1}{V}\frac{\cos\alpha}{\sin\alpha}\left[(M^2-1)\frac{\partial V}{\partial s}\frac{\sin^2\alpha}{\cos\alpha}+\frac{\partial V}{\partial n}\sin\alpha\right]\Delta\sigma$$
$$+\left(\frac{2\omega\cos\alpha}{V}-\frac{\sin\theta\sin\alpha}{r}\right)\Delta\sigma. \tag{14}$$

Finally from Eqs. (12) and (14) it follows that

---

[3] A similar equation for two dimensional flows is found in T. von Kármán, *Compressibility effects in aerodynamics*, J. Aero. Sci. **8**, 337–356 (1941).

$$\Delta\theta = \frac{\Delta V}{V}\cot\alpha\,\frac{(M^2-1)\cdot\tan^2\alpha\cdot\dfrac{\partial V}{\partial s}\cos\alpha+\dfrac{\partial V}{\partial n}\sin\alpha}{\dfrac{\partial V}{\partial s}\cos\alpha+\dfrac{\partial V}{\partial n}\sin\alpha}$$

$$+\left(\frac{2\omega\cos\alpha}{V}-\frac{\sin\theta\sin\alpha}{r}\right)\Delta\sigma. \tag{15}$$

5. The basic equation of the method of characteristics. This follows immediately from Eq. (15). In this case the angle α equals the Mach angle or

$$\tan\alpha=\frac{1}{\sqrt{M^2-1}} \tag{16}$$

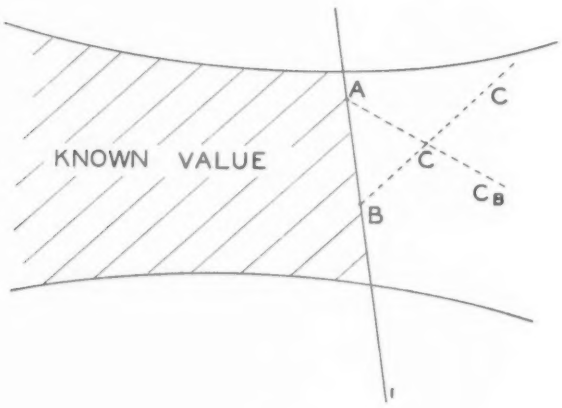


FIG. 4

and so Eq. (15) simplifies to

$$\Delta\theta=\frac{\Delta V}{V}\cot\alpha+\left(\frac{2\omega\cos\alpha}{V}-\frac{\sin\theta\sin\alpha}{r}\right)\Delta\sigma. \tag{17}$$

Thus, it is seen, that moving along characteristics involves an enormous simplification of the differential equations. In the case of two-dimensional, irrotational motion, $\omega=0$, $r=\infty$ and Eq. (17) reduces to the well-known relation

$$\Delta\theta=\frac{\Delta V}{V}\cot\alpha. \tag{18}$$

6. The expression for the rotation. This is needed when the motion is rotational. It is shown[4] that ω is related to the rate of change of the entropy S normal to the streamline:

[4] *Cf.* A Vazsonyi, *On rotational gas flows*, Quart. Appl. Math. **3**, 29–37 (1945).

$$2\omega = \frac{p}{R\rho V}\frac{\partial S}{\partial n}, \tag{19}$$

where S is the specific entropy and R is the gas constant. It can also be expressed simply in terms of the gradient of the total pressure pt as follows:

$$2\omega = -\frac{p}{p_t\rho V}\frac{\partial pt}{\partial n}. \tag{20}$$

7. The essence of the method of characteristics is the following. Suppose the flow is to be determined in the channel of Fig. 4 and the flow is known up to the line l. Select points A and B on line l. At these points the velocity, the direction of flow, etc., is known. Thus characteristics C_a and C_b can be drawn. Equation (17) used along the two different characteristics gives two equations for the velocity and its direction at point C. Thus properties of the flow at C can be determined by solving these simultaneous equations. By repeating this process the complete flow field can be computed or constructed.

CONTINUOUS HEATING OF A HOLLOW CYLINDER*

By G. COMENETZ** (*Geophysical Laboratory, Carnegie Institution of Washington*)

1. Introduction. Formulas are given here for the temperatures within the wall of an infinitely long hollow cylinder which is supplied with heat through the inner surface. The thermal coefficients are assumed constant, the initial temperature of the wall is zero, and the outer surface is either maintained at zero or thermally insulated. Heat is transmitted at any one time uniformly all over the inner surface, but the rate of heat input is permitted to vary with the time, linearly, or at most quadratically. This, of course, is far from a general mode of variation. However, in an application in which a linear or quadratic rate is at least a permissible approximation, the formulas will be of value. They have been useful in connection with the heating of a gun firing steadily, in which, as the barrel grows hot, the heat input through the bore surface decreases nearly enough linearly with the time. Other possible applications are to the heating of a tubular furnace, a chimney, or the insulation on a wire carrying an electric current. It may not be out of place if the writer remarks that in his experience with physics, the occasions on which a first approximation is as much as is required are at least as numerous as those where only an exact theory will do.

The results follow quite directly from formulas given by Carslaw and Jaeger.[1] Certain slowly convergent infinite series arise in the derivations, however, and the main effort below is directed toward summing these series in finite terms. The re-

* Received April 11, 1947. The information described in this article was obtained under contract OEMsr-51 with the Office of Scientific Research and Development, under the supervision of Division One of the National Defense Research Committee.

** Now at Westinghouse Research Laboratories, East Pittsburgh, Pa.

[1] H. S. Carslaw and J. G. Jaeger, *Some two-dimensional problems in conduction of heat with circular symmetry*, Proceedings of the London Mathematical Society (2) **46**, 361–388 (1939) (referred to below as CJ).

sults can be looked on as summation formulas for certain infinite series of orthogonal Bessel functions, and may have separate interest under that aspect. A generalization of one of Carslaw and Jaeger's formulas appears in Sec. 3.

In Sec. 5 we consider briefly the case of a heat input which is not continuous, but consists of a succession of discrete pulses occurring at equal intervals of time. Here again the gun is an example, for heat from each round is transferred to the barrel almost immediately as the round is fired, in a time that is small compared to the time between successive rounds, even at the most rapid machine-gun rate of fire. A definition, resting on a conjecture, is laid down of "equivalent steady input temperature," and an approximate relation is stated between this and the actual temperature produced by the discrete pulses. The considerations of this section relate directly only to the case of a semi-infinite solid. They can, however, be used to make initial approximations in a problem of heat transfer by uniform pulses across a surface of any shape, cylindrical or other, provided the least radius of curvature of the surface is large compared to the distance the temperature peak can travel through the given material in the time between pulses.

The writer wishes to thank Mr. Chester Snow, of the National Bureau of Standards, for some valuable suggestions, particularly in regard to proofs.

2. Outer surface at zero, input quadratic in time. Let the initial temperature of the hollow cylinder be zero, the outer surface be maintained at temperature zero, and heat be supplied across the inner surface, beginning at the initial time $t=0$, at a rate which is expressed by a quadratic polynomial in the time:

$$q(t) = c_0 + c_1 t + c_2 t^2, \tag{1}$$

where c_0, c_1 and c_2 are given constants. The problem is to find the temperature θ throughout the wall as a function of the radius r and the time t, for $a \leqq r \leqq b$ and $t > 0$.

Carslaw and Jaeger derived the following formula[2] for a problem otherwise the same, but where the heat input, instead of being given by (1), consists of a single pulse of strength B at the initial time, $t=0$, the inner surface being insulated for $t>0$. The temperature θ is thus given by

$$\theta = \frac{\pi^2 a B}{2\rho c} \sum_{s=1}^{\infty} e^{-\alpha\beta_s^2 t} A_s(r), \tag{2}$$

where ρc denotes the specific heat (per unit volume), α the diffusivity,

$$A_s(r) = \frac{\beta_s^2 J_1^2(a\beta_s) C(a\beta_s, b\beta_s) C(r\beta_s, b\beta_s)}{J_1^2(a\beta_s) - J_0^2(b\beta_s)}, \tag{3}$$

$\beta_1, \beta_2, \cdots$ are the successive positive roots[3] of

$$J_0(b\beta_s) Y_1(a\beta_s) - Y_0(b\beta_s) J_1(a\beta_s) = 0, \tag{4}$$

and

$$C(x, y) = J_0(x) Y_0(y) - Y_0(x) J_0(y).$$

Let the temperature produced by the B cal/cm^2 entering instantaneously at $t=0$

[2] CJ, p. 385, case (i) with $r'=a$ and $h=0$. Their Q is here $2\pi aB/\rho c$.

[3] Tabulated in Jahnke-Emde, *Table of functions*, Dover Publications, New York, 1943, pp. 207–209.

be written as $Bf(r, t)$. Then the temperature produced by $q(t)$ cal/cm²sec entering continuously beginning at $t=0$ is given by

$$\theta(t) = \int_0^t q(\tau) f(r, t-\tau) d\tau. \tag{5}$$

Applying (5) with $q(t)$ expressed by (1) and $f(r, t)$ by the right-hand member of (2) divided by B, we find that

$$\theta = P(r) + tQ(r) + t^2 R(r) - c_0 S_1(r, t) + c_1 S_2(r, t) - 2c_2 S_3(r, t), \tag{6}$$

where

$$\left.\begin{aligned} P(r) &= c_0 S_{10} - c_1 S_{20} + 2c_2 S_{30}, \\ Q(r) &= c_1 S_{10} - 2c_2 S_{20}, \\ R(r) &= c_2 S_{10}, \end{aligned}\right\} \tag{7}$$

$$S_j(r, t) = \frac{\pi^2 a}{2\rho c} \sum_{s=1}^{\infty} e^{-\alpha\beta_s^2 t} A_s(r)/(\alpha\beta_s^2)^j, \qquad j = 1, 2, 3, \tag{8}$$

$$S_{j0} = S_j(r, 0). \tag{9}$$

Now the first three terms of the right-hand member of (6) satisfy all the conditions imposed on the temperature except the condition that the initial temperature is zero; i.e., they satisfy the Fourier equation

$$\frac{\partial\theta}{\partial t} = \alpha\left(\frac{\partial^2\theta}{\partial r^2} + \frac{1}{r}\frac{\partial\theta}{\partial r}\right) \tag{10}$$

and the boundary conditions

$$\theta(b, t) = 0, \qquad -k\frac{\partial\theta(a, t)}{\partial r} = c_0 + c_1 t + c_2 t^2, \tag{11}$$

where k denotes the conductivity. It is easy to verify this formally by considering large values of t and taking into account that $S_j(r, t)$ approaches zero as t increases indefinitely. It is, therefore, possible to obtain P, Q, R in finite terms, as follows.

Replacing θ in (10) and (11) by $P+tQ+t^2R$ and equating like powers of t, we find that $R(r)$ is completely determined by the conditions

$$(rR')' = 0, \qquad R(b) = 0, \qquad -kR'(a) = c_2. \tag{12}$$

Then $Q(r)$ is determined by

$$(rQ')' = \frac{2Rr}{\alpha}, \qquad Q(b) = 0, \qquad -kQ'(a) = c_1, \tag{13}$$

and finally $P(r)$ by

$$(rP')' = \frac{Qr}{\alpha}, \qquad P(b) = 0, \qquad -kP'(a) = c_0. \tag{14}$$

Entering the resulting expressions for P, Q, R in (7), we derive closed formulas for the series S_{j0}:

$$S_{10} = \frac{a}{k} \ln \frac{b}{r}, \tag{15}$$

$$S_{20} = \frac{-a}{4\alpha k}\left\{[r^2 + a^2(1+2\mu)] \ln \frac{b}{r} + r^2 - b^2\right\}, \tag{16}$$

$$S_{30} = \frac{a}{128\alpha^2 k}\left\{2[r^4 + 4a^2r^2 + 9a^4 - 8a^2b^2 + 8a^2\mu(r^2 + 3a^2 + 2a^2\mu)] \ln \frac{b}{r} + [3r^2 - 5b^2 + 8a^2(1+2\mu)](r^2 - b^2)\right\}, \tag{17}$$

where $\mu = \ln (b/a)$.

The solution to the problem of this section is therefore given by formula (6), with symbols defined by (7), (15), (16), (17), (8) and (3), and the coefficients β determined from (4).

Equation (15) may be written as

$$\sum_{s=1}^{\infty} \frac{J_1^2(a\beta_s)C(a\beta_s, b\beta_s)C(r\beta_s, b\beta_s)}{J_1^2(a\beta_s) - J_0^2(b\beta_s)} = \frac{2}{\pi^2} \ln \frac{b}{r}. \tag{18}$$

This is a summation formula for a particular infinite series of Bessel functions. Equations (16) and (17) are similar formulas. It may be observed that the terms of the series are orthogonal in the interval (a, b) with respect to the weight function r. This is so because the $A_s(r)$ in (3) are a set of characteristic functions for a Sturm-Liouville problem from which (2) may be derived.[4]

As x becomes large,

$$\begin{aligned} J_0(x) \sim -\, Y_1(x) &\sim \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\pi}{4}\right), \\ J_1(x) \sim Y_0(x) &\sim \sqrt{\frac{2}{\pi x}} \sin\left(x - \frac{\pi}{4}\right). \end{aligned} \tag{19}$$

From these well-known approximations, it follows for large s in (3) and (4) that

$$\beta_s \sim \pi(s - \tfrac{1}{2})/(b - a) \tag{20}$$

$$A_s(r) \sim 4(b - r)/\pi^2(ar)^{1/2}(b - a)^2. \tag{21}$$

This makes it clear that the convergence of the series for $S_j(r, 0)$ in (8) is like that of the reciprocal 2nd, 4th, and 6th powers of the integers, and therefore slow. For this reason it was desirable to sum the series in finite terms.

3. Extension of a formula of Carslaw and Jaeger. Consider a hollow cylinder initially at temperature zero. Let B cal/cm^2 of heat be liberated instantaneously at $t=0$ over a cylindrical surface of radius r' in the hollow cylinder. Let the heat radiate to a medium at temperature zero from the inner and outer surfaces, according to the conditions

[4] Cf. H. S. Carslaw, *Mathematical theory of the conduction of heat*, Macmillan, 1921, pp. 127–129. The pulse of heat can be treated as a limiting case of an initial temperature which has a constant value T from $r=a$ to $a+\epsilon$, and zero from $a+\epsilon$ to b, where ϵT remains constants as $\epsilon \to 0$.

$$\frac{\partial\theta}{\partial r} = h\theta \quad \text{at} \quad r = a, \qquad \frac{\partial\theta}{\partial r} = -h'\theta \quad \text{at} \quad r = b, \tag{22}$$

where h, h' are constants not both zero. Then

$$\theta = \frac{\pi^2 r' B}{2\rho c} \sum_{s=1}^{\infty} e^{-a\beta_s^2 t} \frac{\beta_s^2[\beta_s J_1(a\beta_s) + hJ_0(a\beta_s)]^2 F(r)F(r')}{(h'^2+\beta_s^2)[\beta_s J_1(a\beta_s)+hJ_0(a\beta_s)]^2 - (h^2+\beta_s^2)[\beta_s J_1(b\beta_s) - h'J_0(b\beta_s)]^2}, \tag{23}$$

where

$$F(x) = J_0(x\beta_s)[-\beta_s Y_1(b\beta_s) + h'Y_0(b\beta_s)] + Y_0(x\beta_s)[\beta_s J_1(b\beta_s) - h'J_0(b\beta_s)] \tag{24}$$

and $\beta_1, \beta_2, \cdots$ are the successive positive roots of

$$[\beta_s J_1(a\beta_s) + hJ_0(a\beta_s)][\beta_s Y_1(b\beta_s) - h'Y_0(b\beta_s)] - [\beta_s J_1(b\beta_s) - h'J_0(b\beta_s)][\beta_s Y_1(a\beta_s) + hY_0(a\beta_s)] = 0. \tag{25}$$

If h and h' are both zero the term $2r'B/\rho c(b^2-a^2)$ must be added to the right member of (23).

Carslaw and Jaeger[5] give this formula for the case where $h = h'$. The more general case above, which was derived by their method, would be required if the problem of the next section were to be generalized to permit radiation at the outer surface.

In particular, if the outer surface is thermally insulated, and the heat is liberated at the inner surface with that surface thereafter thermally insulated ($h = h' = 0$, $r' = a$), then

$$\theta = BE_0 + \frac{\pi B}{\rho c} \sum_{s=1}^{\infty} e^{-a\beta_s^2 t} E_s(r), \tag{26}$$

where

$$E_0 = \frac{2a}{\rho c(b^2 - a^2)}, \qquad E_s(r) = \frac{\beta_s J_1^2(b\beta_s)[J_1(a\beta_s)Y_0(r\beta_s) - Y_1(a\beta_s)J_0(r\beta_s)]}{J_1^2(a\beta_s) - J_1^2(b\beta_s)} \tag{27}$$

and $\beta_1, \beta_2, \cdots$ are the successive positive roots[6] of

$$J_1(a\beta_s)Y_1(b\beta_s) - Y_1(a\beta_s)J_1(b\beta_s) = 0. \tag{28}$$

The reduction of (23) to (26) depends on the use of (28) and the identity

$$Y_0(x)J_1(x) - Y_1(x)J_0(x) = \frac{2}{\pi x}. \tag{29}$$

4. Outer surface insulated, input linear in time. Let the initial temperature of a hollow cylinder be zero, the outer surface be thermally insulated, and heat be supplied across the inner surface beginning at $t = 0$ at a rate linear in the time:

$$q(t) = c_0 + c_1 t. \tag{30}$$

---

[5] CJ, p. 385, case (iii). They omit the additional term for $h = h' = 0$. It enters because the integrand in the inversion formula has a simple pole at the origin with this term as residue. Of course, the term is physically necessary.

[6] Tabulated in Jahnke-Emde, *loc. cit.*, pp. 204–206.

The problem is to find the subsequent temperature $\theta(r, t)$. The q is taken as merely linear here because the resulting formulas are about as involved as those for a quadratic q in the problem of Sec. 2.

As in Sec. 2, we apply (5) with $q(t)$ given by (30) and $f(r, t)$ by the right member of (26) divided by B. The result is.

$$\theta = E_0(c_0 t + \tfrac{1}{2}c_1 t^2) + P(r) + tQ(r) - c_0 S_1(r, t) + c_1 S_2(r, t), \tag{31}$$

where

$$P(r) = c_0 S_{10} - c_1 S_{20}, \qquad Q(r) = c_1 S_{10}, \tag{32}$$

$$S_j(r, t) = \frac{\pi}{\rho c} \sum_{s=1}^{\infty} e^{-\alpha\beta_s^2 t} E_s(r)/(\alpha\beta_s^2)^j, \tag{33}$$

$$S_{j0} = S_j(r, 0), \qquad j = 1, 2. \tag{34}$$

Proceeding as before, we recognize that the first three terms of the right member of (31) obey (10) with the conditions

$$\frac{\partial\theta(b, t)}{\partial r} = 0, \qquad -k\frac{\partial\theta(a, t)}{\partial r} = c_0 + c_1 t. \tag{35}$$

These lead to the conditions

$$(rQ')' = \frac{E_0 c_1 r}{\alpha}, \qquad Q'(b) = 0, \qquad -kQ'(a) = c_1, \tag{36}$$

$$(rP')' = \frac{r}{\alpha}(E_0 c_0 + Q), \qquad P'(b) = 0, \qquad -kP'(a) = c_0. \tag{37}$$

However, in solving (36) for $Q(r)$ a new difficulty arises, for it develops that one of the two boundary conditions is redundant in view of the definition of E_0 in (27) and the relation $\rho c\alpha = k$. The consequence is that after full use has been made of (36) and (37), a constant of integration remains undetermined in $P(r)$. To fix its value another condition is required. We may find one as follows.

Since the outer surface is insulated, all the heat that has entered the cylinder up to any time t is still there, or in symbols,

$$\int_0^t 2\pi a(c_0 + c_1 t)dt = \int_a^b 2\pi r\rho c\theta(r, t)dr. \tag{38}$$

Here again, taking t large, we recognize that the relation must hold for the first three terms of the right member of (31). Substituting these terms for $\theta(r, t)$ in (38) and using the definition of E_0, we arrive at the conditions

$$\int_a^b rQdr = 0, \qquad \int_a^b rPdr = 0. \tag{39}$$

The first is again redundant, but the second serves to fix the constant in $P(r)$.

With P and Q thus determined, the S_{j0} can be found from (32):

$$S_{10} = \frac{E_0}{8\alpha}\left(4b^2 \ln\frac{b}{r} + 2r^2 + 2\mu\nu - a^2 - 3b^2\right). \tag{40}$$

$$S_{20} = \frac{a}{16\alpha k(b^2 - a^2)} \{4b^2r^2 \ln r - \nu(b^2 - a^2 - 4\mu b^2) \ln r$$
$$- \tfrac{1}{2}r^4 - (b^2 - a^2 + 4b^2 \ln b + 2\mu\nu)r^2$$
$$+ \nu[(a^2 + 3b^2)\mu + (b^2 - a^2 - 4\mu b^2) \ln b - 2\mu^2\nu] - \tfrac{1}{6}(2a^4 + 2a^2b^2 - 7b^4)\}, \quad (41)$$

where $\nu = 2a^2b^2/(b^2 - a^2)$.

The solution of the problem of this section is given by (31), with symbols defined by (32), (33), (40), (41), and (27), and the β's defined by (28).

Equation (40) may be written

$$\sum_{s=1}^{\infty} E_s(r)/\beta_s^2 = \frac{a}{4\pi(b^2 - a^2)}\left(4b^2 \ln \frac{b}{r} + 2r^2 + 2\mu\nu - a^2 - 3b^2\right) \quad (42)$$

and similarly for (41). Here again the $E_s(r)$ may be shown to be orthogonal in the interval (a, b) with respect to the weight function r. In this case a constant function is needed to complete the orthogonal set, as appears from the associated Sturm-Liouville problem.

As before, it can be shown from (19) that the series for S_{10} and S_{20} converge slowly, like the reciprocal 2nd and 4th powers of the integers.

5. Intermittent Heating. Consider a semi-infinite solid of constant thermal coefficients, initially at temperature zero, which receives heat through its surface in a series of equal pulses of strength B cal/cm^2. These are to occur at equal intervals of time, with a frequency of n times per second, the first at $t = 0$; and the surface is to be thermally insulated between pulses. The resulting temperatures have then a fluctuating character near the surface, but deeper in the solid the fluctuations become damped out, and the temperatures approximate more nearly to what would be produced by a continuous constant input of heat through the surface. The constant rate of what may be called the "equivalent steady input" is of course nB cal/cm^2 sec. But the best approximation to the fluctuating temperature is obtained by making the equivalent steady input begin, not at $t = 0$, but at $t = -1/2n$; that is, half the time of a pulse before the first pulse. The last statement is not one that the writer has been able to prove, although on the basis of numerical computation he is certain that it is at least very nearly true. Physically, the equivalent steady input temperature is what an instrument situated at a depth in the solid would report the temperature to be, if the instrument were not quick or sensitive enough to follow the individual fluctuations, but only their average in time. It is thus clear that the concept of equivalent steady input may be important in intermittent heating, and it would be of interest if the conjecture above could be proved.

The actual fluctuating temperature and the equivalent steady input temperature differ most at the surface of the semi-infinite solid, but there is a simple relation between them. It can be shown that at the instant before the $(m+1)$st pulse, the equivalent steady input temperature (input commencing at $t = -1/2n$) exceeds the fluctuating temperature at the surface by almost exactly

$$1.46B(n/\pi\rho ck)^{1/2} \quad (43)$$

after the first three or four pulses. We omit the simple proof, which depends on standard semi-infinite solid heating formulas and Euler's summation formula. The

value of this relation is that, together with a steady input calculation, it permits an estimate to be made of the base temperature level at the surface, upon which the effect of the $(m+1)$st pulse is superimposed. The character of the $(m+1)$st temperature surge can thus be calculated without the necessity of summing the separate effects of the m preceding pulses.

NOTE ON RAYLEIGH'S METHOD AND THE NON-UNIFORM STRUT*

By H. A. LANG (*Cornell University*)

For a strut of the non-uniform flexural rigidity $B=EI(x)$ under end thrust,[1] Rayleigh's method consists of estimating the lowest critical load P_1 by using an assumed deflection y and computing the expression

$$P = \frac{\int_0^L B(y'')^2 dx}{\int_0^L (y')^2 dx}. \tag{1}$$

If P_1'' denotes the lowest value of P obtained for an assumed deflection y which satisfies the end conditions, it can be proved that $P_1' > P_1$, i.e., Rayleigh's method always leads to an overestimate.

An equivalent procedure may be based on the relation

$$P = \frac{\int_0^L (y')^2 dx}{\int_0^L \left(\frac{y^2}{B}\right) dx}. \tag{2}$$

Equation (2) is obtained by eliminating y'' from (1) through use of the governing differential equation $By''+Py=0$. The lowest critical load obtained from (2) for an assumed deflection y is denoted by P_1'.

It is commonly taken for granted that (2) yields (a) an overestimate, and (b) a lower value than (1). These statements, however, are not obvious consequences of the fact that (1) gives an overestimate and no proof of them seems to have been given. A simple proof is presented here, i.e., it is shown that $P_1'' > P_1' > P_1$. The deflection y is expanded in the form $y=\sum_1^\infty A_N Y_N(x)$ where $Y_N(x)$ is the deflection function corresponding to the nth mode and so satisfies $BY_N''+P_N Y_N=0$. Here P_n is the nth critical load. The functions Y_n are supposed ordered so that Y_1 and P_1 correspond to the first mode.

To simplify the discussion let

* Received April 10, 1947.

[1] For a detailed discussion see R. V. Southwell, *Theory of elasticity*, Oxford, 1946, pp. 446–453.

$$I_N = A_N^2 \int_0^L [Y_N']^2 dx.$$

The required integrals are then:

$$\int_0^L (y')^2 dx = \sum_1^\infty I_N, \qquad \int_0^L \left(\frac{y^2}{B}\right) dx = \sum_1^\infty \frac{I_N}{P_N}, \qquad \int_0^L B(y'')^2 dx = \sum_1^\infty P_N I_N.$$

The integrals depend on the relations

$$\int_0^L (Y_M')(Y_N') dx = \int_0^L \frac{(Y_M)(Y_N)}{B} dx = \int_0^L B(Y_M'')(Y_N'') dx = 0, \quad [m \neq n]$$

$$\int_0^L B[Y_N'']^2 dx = P_N \int_0^L [Y_N']^2 dx, \qquad \int_0^L \frac{[Y_N]^2}{B} dx = \frac{1}{P_N} \int_0^L [Y_N']^2 dx,$$

the details of which are presented by R. V. Southwell[1] when the strut is pinned. Other common end conditions are accounted for by imagining the strut as extended so that the pinned ends correspond to two adjacent intersections of the centerline of the strut with the line of thrust.

To establish the inequalities, we first write (2) in the form

$$P_1' = \frac{I_1 + I_2 + \cdots}{(I_1/P_1) + (I_2/P_2) + \cdots} = P_1 \frac{I_1 + I_2 + \cdots}{I_1 + I_2(P_1/P_2) + \cdots}.$$

Then since $P_1/P_N < 1$ for each integer $N \geqq 2$ the term in brackets exceeds unity and so $P_1' > P_1$.

To establish the inequality $P_1'' > P_1'$ it is sufficient to show that

$$\frac{P_1 I_1 + P_2 I_2 + \cdots}{I_1 + I_2 + \cdots} > \frac{I_1 + I_2 + \cdots}{(I_1/P_1) + (I_2/P_2) + \cdots}.$$

But this is equivalent to

$$[P_1 I_1 + P_2 I_2 + \cdots]\left[\frac{I_1}{P_1} + \frac{I_2}{P_2} + \cdots\right] > [I_1 + I_2 + \cdots].^2$$

For each m and n corresponding cross terms are:

$$I_M I_N \left[\frac{P_N}{P_M} + \frac{P_M}{P_N}\right] \quad \text{and} \quad 2I_M I_N, \qquad (m \neq n).$$

When $m = n$ the terms of form I_N^2 which appear on each side of the inequality are equal.

The inequality is established if for each m and n

$$\left[\frac{P_N}{P_M} + \frac{P_M}{P_N}\right] > 2 \qquad (m \neq n),$$

but multiplying by $P_n P_m$ this reduces to $(P_N - P_M)^2 > 0$ and the result follows.[2]

[2] Professor G. F. Carrier called the author's attention to the fact that this technique was applied to vibration problems in the paper *Ueber das Gegenstueck zum Rayleighschen Verfahren der Schwingungslehre* by K. Hohenemser and W. Prager, Ing.-Arch., **3**, 306–310 (1932).

BOOK REVIEWS

Six-place tables. By E. S. Allen. McGraw-Hill Book Company, Inc., New York and London, 7th ed., 1947. xxiii+232 pp. $2.50.

In this seventh edition the tables of the natural logarithm, exponential and hyperbolic functions have been extended to six places, the argument being given to four, three, and three places, respectively. For numbers less than unity $\ln x+10$ is given instead of $-\ln x$ as in the previous edition.

S. PRAGER

Tables of integrals and other mathematical data. By H. B. Dwight. Macmillan Company, New York, 2nd ed., 1947. viii+250 pp. $2.50.

In this revised edition additional material has been included on various progressions, powers of Taylor series, roots of quadratic equations, square roots of complex quantities, inverse trigonometric functions, probability integrals, hyperbolic and Bessel functions. New indefinite integrals involving $a+bx$, $(ax^2+bx+c)^{1/2}$, $\sin x$, $\cos x$, and Bessel functions have also been added, together with some definite integrals. Further additions include a table of the normal probability integral and an extension of the bibliography.

S. PRAGER

Tables of supersonic flow around cones. By the staff of the Computing Section, Center of Analysis, Massachusetts Institute of Technology, Department of Electrical Engineering. Under the direction of Zdenek Kopal. Cambridge, Massachusetts, 1947. xviii+555 pp. $5.00.

The introduction to this volume is a concise review of the derivation of the ordinary non-linear differential equation defining the supersonic flow of a compressible fluid past a circular cone and a statement of the relations among the state variables on either side of the shock implied by such a flow. The remainder of the book is a tabulation of the numerically obtained solutions of this problem.

For air ($\gamma=1.405$), the velocity components and velocity of sound are tabulated against the flow direction parameter θ for cone angles of 5°, 7.5, · · · , 22.5, 25, 30, 35, 40, 45, and 50°. For each cone a complete range of wave angles is considered. The increments in θ vary from 1/4° to 1° for various solutions and regions. In general there is five digit accuracy. A less extensive but similar table is presented for $\gamma=4/3$. Several cross tabulations are also included.

G. F. CARRIER

Gas dynamics tables for air. By Howard W. Emmons. Dover Publications, Inc., New York, 1947. 46 pp. $1.75.

Dr. Emmons has compiled tables of the most commonly used functions occurring in gas dynamics. The quantities tabulated include pressure, velocity, density, stream-tube area, dynamic pressure, temperature, velocity of sound and Prandtl-Meyer angles, as functions of Mach number over the complete range. Similar quantities following a normal shock wave are also tabulated.

A sample of five hundred of the numbers in these tables was taken to check the computing accuracy, and it was found that there were four errors in the fifth published figure and two errors by one unit in the fourth figure. It appears then that, in general, the first four figures are accurate. Unpublished tables similar to these have been in use by the reviewer for several years and have been found to be of great practical utility in performing computations involving the flow of room-temperature air. These tables are accordingly highly recommended for the use of aerodynamicists.

ARTHUR KANTROWITZ

Reviews, each about 500 words in length, will be written and signed by authorities in the particular fields of mechanics involved. Moreover, every effort will be made to keep the lapse of time between the publication date of the paper and its review to less than two months, thereby making APPLIED MECHANICS REVIEWS a truly "NEWS" Journal.

to keep in touch with the work of your colleagues in the research centers of the ... with a minimum expenditure of time and effort, don't fail to subscribe for AP- MECHANICS REVIEWS.

... of one of the Associations cooperating with the A.S.M.E. in the publication of this ... you may take advantage of the Special Charter Rate of $9.00 (Regular price $12.50 an-

... the first number and the succeeding issues for 1948—mail your subscription ... to—

... SOCIETY OF MECHANICAL ENGINEERS
... West 39th Street, New York 18, N.Y.

...S OF NOMOGRAPHY

... D. DOUGLASS and DOUGLAS P. ADAMS, Massachusetts Institute of Technology. 213 pages, $3.50

...understanding, design, erection, and practical use of the alignment diagram, with emphasis on the mathematical foundation of the nomogram theory.

...LIED MATHEMATICS FOR ENGINEERS AND PHYSICISTS

...LOUIS A. PIPES, formerly of Harvard University. 624 pages, $5.50

...topics of higher mathematics related to electrical, mechanical, and civil engineering as well as the methodology of classical physics.

...EMATICAL METHODS IN ENGINEERING

...THEODORE V. KÁRMÁN, California Institute of Technology, and MAURICE A. BIOT, Columbia University. 340 pages, $4.50

...to familiarize engineering students with the special techniques used in the mathematical treatment of engineering problems.

...K OF MATHEMATICAL TABLES

...A. FLETCHER, J. C. P. MILLER, and L. ROSENHEAD, University of Liverpool. 451 pages, 1960

...index to all published and some unpublished mathematical tables, compiled by ... of the foremost recognized experts in the international field of mathematical tables and ...ability.

Send for copies on approval

McGRAW-HILL BOOK COMPANY, Inc.
New York 18, N.Y.
330 West 42nd Street